

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:56:23 ; Search time 37.5 seconds
(without alignments)
10481.029 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattccattgtgtgggta.....tgggattcctaaaaaaaaaaa 2043

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09645078/runat_06052004_104600_21755/app_query.fasta_1.2183
-DB=PIR_78 -QMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09645078.cgn_1.1.52/runat_06052004_104600_21755 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	628.5	16.8	484	JC7350	N-acetylglucosamin
2	615.5	16.4	486	JC7351	N-acetylglucosamin
3	588.5	15.7	484	JE0261	chondroitin 6-sulf
4	500.5	13.4	458	A57397	probable enzyme, C
5	168	4.5	307	E95934	BHLFI protein - hu
C 6	147	4.0	660	Q0B23	collagen alpha 1(I)
C 7	147	4.0	1433	S21826	collagen alpha 1(I)
C 8	145	3.9	1464	1 CGHULS	collagen alpha 1(I)
C 9	143.5	3.8	1172	1 TSHUP2	thrombospondin 2 p
C 10	140.5	3.8	1464	2 S59856	collagen alpha 1(I)
C 11	138.5	3.7	940	2 JE0291	FB19 protein - hum
C 12	138.5	3.7	1373	1 A43291	collagen alpha 2(I)
C 13	138.5	3.7	1487	1 CGHUI6	collagen alpha 1(I)
C 14	138	3.7	319	2 F75420	hypothetical prote

C 15	138	3.7	671	1 CGRTL5	collagen alpha 1(I)
C 16	137.5	3.7	1496	1 CGHU2V	collagen alpha 2(V)
C 17	136	3.7	1049	1 CGBO7S	collagen alpha 1(I)
C 18	136	3.7	1964	2 T09059	notch4 - mouse
C 19	135	3.6	784	2 JQ0317	hypothetical 82x p
C 20	134.5	3.6	1418	2 T45467	collagen alpha 1(I)
C 21	134	3.6	337	2 T23794	hypothetical prote
C 22	134	3.6	1042	1 CGCH1S	collagen alpha 1(I)
C 23	133.5	3.6	1419	2 A41182	collagen alpha 1(I)
C 24	133.5	3.6	1487	2 B41182	collagen alpha 1(I)
C 25	133	3.6	825	1 EDBEXD	immediate-early pr
C 26	133	3.6	1215	2 T32734	myosin-1A - Acanth
C 27	133	3.6	1360	2 T33922	hypothetical prote
C 28	132	3.5	286	2 S34665	collagen, cuticula
C 29	132	3.5	325	2 T32248	hypothetical prote
C 30	131	3.5	636	2 S41067	collagen alpha 1(I)
C 31	131	3.5	2944	2 A54849	collagen alpha 1(V)
C 32	130.5	3.5	316	2 T19288	hypothetical prote
C 33	130.5	3.5	1122	2 T14180	LRG5 protein - Chl
C 34	130.5	3.5	1700	2 S08167	collagen alpha 2(I)
C 35	129.5	3.5	1414	1 S23809	collagen alpha 1(I)
C 36	128	3.5	632	2 S42731	collagen alpha 1 c
C 37	127.5	3.4	1184	2 S50832	atrophin-1 - human
C 38	127	3.4	1106	2 JQ0405	hypothetical 119.5
C 39	126	3.4	1366	1 CGHU2S	collagen alpha 2(I)
C 40	126	3.4	1712	1 CGHU2B	collagen alpha 2(I)
C 41	125.5	3.4	316	2 T19291	hypothetical prote
C 42	125.5	3.4	640	2 T08179	collagen alpha 1(X)
C 43	125	3.4	680	1 CGHUID	collagen alpha 1(I)
C 44	125	3.4	1466	1 CGHU7L	collagen - sea urc
C 45	124.5	3.4	290	2 A32249	

ALIGNMENTS

RESULT 1

JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7350
R:Uchimura, K.; Pasakany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; sulfotransferase; sulfotransferase protein

Alignment Scores:
Pred. No.: 2.22e-39 Length: 484
Score: 628.50 Matches: 154
Percent Similarity: 48.28% Conservative: 71
Best Local Similarity: 33.05% Mismatches: 177
Query Match: 16.80% Indels: 64
DB: 2 Gaps: 11

US-09-645-078-1 (1-2043) x JC7350 (1-484)

QY	19	TACGGAAGACGACAGAGGGGTAGAGGAGAGAAAGCGCATGGCCCGG---CTAGCAGTG---72
Db	35	HisserGluGlnAspLysGlyArgAsnCysProGlyLeuGlnArgSerLeuGlyValTrp 54
QY	73	AGCTCTCAAAGCAGCAGGAGCCCAAGCCCAAGAGTCTTCACATTCAGCACAAATGCT 132
Db	55	SerLeuGluAlaAlaAlaAlaGlyGluArgGluGluAlaGluValArgSerLeuAla 74
QY	133	ACTGCCTAABAAATGAAGCTCTCTGTTTCCAGATGCCCATCTTGGTCT 192

Db	75	-----GluglyAsnProAspArgSerProGlySer	84
QY	193	ATTCTTCCACATGTACAGCCACAACATCAGCTCCCTGTCTATCAAGGCACAGCCGAGG	252
Db	85	-----Pro-GlyAsnLeuSerAlaValGlyGluAlaValThrGlnGlnIu	99
QY	253	CATGCACGCTGGTCTCTCTCTCCCTGGCGCTGTGCTCTCTTTTGTGGGAGAGCTTTT	312
Db	99	sglnhisIleTyrValHisAlaThrTrpArgThrGlySerSerPheLeuGlyGlnLeuP	119
QY	313	TGGCGAGCACCCAGATGTTTTCTACCTGATGAGCCCGCTCGCAGCTGTGATGACCTT	372
Db	119	eAsnGlnHisProAspValPheTyrLeuTyrGluProMetTrpHisLeuTrpGlnAlaLe	139
QY	373	CAAGCAGACCGCTCGATCGTGCACATGCTGTGCGGATCTGATACGGCCGCTCT	432
Db	139	uTyrProGlyAspAlaGluSerLeuGlnGlyAlaLeuArgAspMetLeuArgSerLeuP	159
QY	433	CTTGTGCGCATGACGCTCTTTGATCGCTPACATGGAACCTGTGTCCCGGAGACAG	487
Db	159	eArgCysAspPheSerValLeuArgLeuTyrAlaGlnProGlyAspProGlyGluArgAl	179
QY	488	-----TCCAGCTCTTTTCAGTGGGAGAACAGCCGGCCCT	522
Db	179	aProAspSerAlaAsnLeuThrThrAlaMetLeuPheArgTrpArgThrAsnHisValI	199
QY	523	GTGTTCTGCACCTGCTGTGACATCATCCCAAGATGAA	570
Db	199	eCysSerProProLeuCysProAlaAlaProArgAlaArgAlaAspValGlyLeuValG	219
QY	571	CCGGGCTCACTGCAGGCTCTGTGAGTCAACAGCCCTTTGAGGTGGTGGAGAGCCGTG	630
Db	219	uAspLysAlaCysGluSerThrCysProProValSerLeuArgAlaLeuGluAlaGluCy	239
QY	631	CGCTCTTACAGCCACGCTGGTCTCAAGAGGTGGCTTCTTCAACCTGCAGTCCCTCTA	690
Db	239	sArgLysTyrProValValIleLysAspValArgLeuLeuAspLeuGlyValLeuVa	259
QY	691	CCGCTGTGTGAAGACCCCTCCCTCAACCTGCATCTGTCACCTGTGACCTGGTCCGGACCCCG	750
Db	259	lProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheArgAspProAr	279
QY	751	GGCGTGTTCGCTCCCGAGACGCACAAGGAGATCTCATGATGACAGT	802
Db	279	gAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgGluSerIleGlnVa	299
QY	803	-----CGCATTTGTGTG	814
Db	299	lLeuArgThrArgGlnArgGlyAspHisPheHisArgValLeuLeuAlaHisGlyValAs	319
QY	815	-----GGCGAGCATGAGCAAAACTCAAGAGGAGACCAACCTACTACTAT	859
Db	319	pAlaArgProGlyGlyGlnAlaArgAlaLeuProSerAlaProArgAlaAspPhePheLe	339
QY	860	-----GTGATGACAGGTCACTGCAAGACGAGCTGGAGATCTACAAGACCATCCAGTC	912
Db	339	uThrSerAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuPheThrArgGgl	359
QY	913	CTTGCCCAAGCCCTGCAGGAACGCTACCTCTGTGCGCTATGAGACCTGGGCTCGAGC	972
Db	359	yAlaProAlaTrpLeuArgAArgTyrLeuArgLeuArgTyrGluAspLeuValTrpGl	379
QY	973	CCCTGTGGCCAGACTTCCGAAATGTATGAATTCGTGGGATTTGAAATCTTGCCCATCT	1032
Db	379	nProGlnAlaGlnLeuArgAArgLeuArgPheSerGlyLeuArgThrLeuAlaAlaLe	399
QY	1033	TCAGACTGGGTGATACATCAACCCGAGGCAAGGGCATGGGT	1089
Db	399	uAspAlaPheAlaPheAsnMetThrArgGlySerAlaTyrGlyAlaAspArgProPheHi	419
QY	1090	CACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGGTCTTTGCCCTATGA	1149
Db	419	sLeuSerAlaArgAspAlaArgGluAlaValHisValTrpArgGluArgLeuSerGlnGl	439

QY 1150 AAAGGTTTCTCGACTTCAGAAACCTGTGGCATGCCATGAATTGCTGGTGCTACGCCCA 1209
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 439 uGLNVaRGlnValGluThrAlaCysAlaProAlaMetArgLeuLeuAlaIyrProAr 459
QY 1210 CGTCAGATCTGAACAAGAAGACAGAAAACCTGTGTG-----CTGATCTTCT 1254
Db :|||||.....:|||||.....:|||||.....:|||||.....
QY 459 gSerGlyAspGluArgAspArgLysThrValArgGluGlyGluThrProLeuGluThryl 479
DQ 1255 GTCTACCTGGACTGTC 1270
Db :||:|||||.....:|||||.....:|||||.....:|||||.....
DQ 479 sAlaAsnTrpAlaVal 484

RESULT 2
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
RJuChimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A>Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: Glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Alignment Scores:
Pred. No.: 2,14e-38 Length: 486
Score: 615.50 Matches: 155
Percent Similarity: 48.22% Consistative: 62
Best Local Similarity: 34.44% Mismatches: 163
Query Match: 16.45% Indels: 70
DB: 2 Gaps: 12

US-09-645-078-1 (1-2043) x JC7351 (1-486)

QY 7 CATTGTGTTGGGTACGAGACGACAGAGGCTAGAGGNGAAGAGCGCATGCCTCCGGCTA 66
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 44 HisCyPrGlyLeuGlnArg-----SerLeu 52
QY 67 GCAGTG---AGCTCTCAAAGCAGCAGGAGAACCCACCACAGGCTCTTCCACTTCAG 123
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 53 GlyValTrpSerLeuGluAlaAlaAlaGlyGluArgGluGlnGlyAlaGluAlaArg 72
QY 124 CACAATGCTACTCCCTMAAAATGAAGCTCTGCTGTGTTCTGGTTTCCCAGATGGCCAT 183
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 73 AlaAlaGluGluGlyAlaAsnGlnSerProArg-----PheProSerAsnLeu 89
QY 184 CTGGGCTCTATTCTTCCACATCATACAGCCACAAACATCAGTCCCTGTCTATGAAGGCACA 243
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 90 SerGlyAlaValGlyGluAlaVal-SerArg----- 99
QY 244 GCCCAGCGCATGCAGGCTCTGTTCTCTCTCTGGCGCTGCTGCTCTTCTTTTGGG 303
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 100 ---GluLysGlnHisIleTyrrValHisAlaThrTrpArgInrGlySerPheLeuGl 118
QY 304 GCAGCTTTTTGGGACGACCCACAGATGTTTTTCTACTGATGGAGCCCGCTCGCACGTGTG 363
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 118 yGluLeuPheAsnGlnHisProAspValPheTyrlauTyrrGluProMetTrpHisLeuTr 138
QY 364 GATGACCTTCAAGCAGACGACCGCTGATGCTGCACATGCTGTGGGGATCTGATACG 423
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 138 pGlnAlaLeuTyrrProGlyAspAlaGluSerLeuGlnGlyAlaLeuArgAspMetLeuAr 158
QY 424 GGCGGCTTTCTTGTGGCAGATGAGCGTCTTTTGATGCTCTACATGAAGAACTGGT---CCCCG 480
Db :|||||.....:|||||.....:|||||.....:|||||.....
DQ 158 gSerLeuPheArgCysAspPheSerValLeuArgLeuTyrrAlaProGlyAspProAl 178

```
QY 481 GAGACAG-----TCCAGCCTCTTTCAGTGGGAGACAG 513
Db 178 aAlaArgAlaProAspThrAlaAsnLeuThrAlaAlaLeuPheArgTrpArgThrAs 198
QY 514 CCGGGCCCTGGTTCTCCACCTGCTGCTGATCATCCCAAGATGAA-----562
Db 198 nLysValIleCysSerProProLeuCysProGlyAlaProAlaAlaGluValG1 218
QY 563 -ATCATCCCCGGGCTCAGTCAGGCTCCTGTGCAGCTCAACAGCCCTTTCAGTGGTGA 621
Db 218 yLeuValGluAspThrAlaCysGluArgSerCysProProValAlaIleAlaGluG1 238
QY 622 GAAGCCCTGGCGCTCCCTACAGCCACGCTGGTGTCTCAAGGAGTGGCTTCTCAACCTGCA 681
Db 238 uAlaGluCysArgIleCysTrpProValValIleLysAspValArgLeuLeuAspLeuG1 258
QY 682 GTCCCTTACCGCTGTGAAAGACCCCTCCCTCACTGCATATCATGCACTGTGCTCG 741
Db 258 yValLeuValProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheAr 278
QY 742 GGACCCCGGGCGCTGTTCCGTTCCCGAAGACGCACAAAGGAGATCTCATGATTGACAG 801
Db 278 sAspProArgAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgGluSe 298
QY 802 T-----CGCATTTGTGATG-----814
Db 298 rIleGlnValLeuArgThrArgGlnArgGlyAspArgPheHisArgValLeuLeuAlaHi 318
QY 815 -----GGGAGCATGAGCAAAACTCAAGAGGAGGACCAACCCCTA 855
Db 318 sGlyValGlyAlaArgProGlyGlnSerArgAlaLeuProAlaAlaProAlaAs 338
QY 856 CTAT-----GTGATGAGGTCACTGTGCCAAAGCCAGCTGGAGATCTACAAGAC 903
Db 338 pPhePheLeuThrGlyAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuLeuPh 358
QY 904 CATCAGCTCTTGGCCAGCCCTGCGAGGACGCTACTGCTGTGCGCTATGAGACCT 963
Db 358 eAlaArgGlyAlaProAlaTrpLeuArgArgGlyTrpLeuArgLeuArgGlyAspLe 378
QY 964 GGCTCGAGCCCTGTGGCCAGACCTCCGAATATATGAATTCGTGGGATTTGAATCTT 1023
Db 378 uValArgGlnProArgAlaGlnLeuArgArgLeuLeuArgPheSerGlyLeuArgAlaLe 398
QY 1024 CCCCATTCTCAGACCTGGGTGCATATACATCCCGAGGAGGAGGATGGGT---GACCA 1080
Db 398 uAlaAlaLeuAspAlaPheAlaLeuAsnMetThrArgGlyAlaAlaTrpGlyAlaAspAr 418
QY 1081 CGCTTTCCACAAATGCCAGGATGCCCTTAATGCTCCAGGCTTGGCGCTGCTTT 1140
Db 418 gProPheHisLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrpArgGluArgLe 438
QY 1141 GCCCTATGAAGTTTCTCGACTTCAGAAAGCCTGTGGCGATGCCATGAATTTGCTGGG 1200
Db 438 uSerArgGluGlnValArgGlnValGluAlaAlaCysAlaProAlaMetArgLeuLeuAl 458
QY 1201 CTACGGCCAGCTCAGATCTGAACAAGAA 1228
Db 458 aTrp-----ProArgSerGlyGluGlu 465
```

RESULT 3

JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C:Accession: JE0261
R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.
J. Biochem. 124, 670-678, 1998
A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
A:Reference number: JE0261; MUID:98391845; PMID:9722682
A:Accession: JE0261

A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfate from 3'-phosphoaden-
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Alignment Scores:
Pred. NO.: 2,34e-36 Length: 484
Score: 588.50 Matches: 131
Percent Similarity: 54.40% Conservative: 67
Best Local Similarity: 35.99% Mismatches: 137
Query Match: 15.73% Indels: 29
DB: 2 Gaps: 6

US-09-645-078-1 (1-2043) x JE0261 (1-484)

```
QY 248 GAGCQCATGCAGTCTGCTGCTTCTTCCTTCCTGGGCTCTGCTCTTCTTTGTGGGCGAG 307
Db 117 LysArgHisTrpMetTyrrValPheThrTrpArgSerGlySerSerPhePheGlyGlu 136
QY 308 CTTTGTGGGCGACCCACAGATGTTTCTACCTGATGGAGCCGCTGCGACGTGTGGATG 367
Db 137 LeuPheAsnGlnAsnProGluValPheLeuTyrrGluProValTrpHisValTrpGln 156
QY 368 ACCTTCAAGACGAGCACCCCTGGATGCTGCACATGGCTGTGCGGATCTCATAGGGCC 427
Db 157 LysLeuTyrrProGlyAspAlaValSerLeuGlnGlyAlaAlaArgAspMetLeuSerAla 176
QY 428 GTCTTCTTGTGGCAGCATGAGCTCTTTCATGCTCATCTACCTGGAACCTGTGCTCCCGGAGACAG 487
Db 177 LeuTyrrArgCysAspLeuSerValPheGlnLeuTyrrSerProAlaGlySerGlyGlyArg 196
QY 488 TCC-----AGCCTCTTTCAGTGGGAGAACAGCCGGGCCCTGTGTCTTCTGCACCT 535
Db 197 AsnLeuThrLeuGlyIlePheGlyAlaAlaThrAsnLysValValCysSerSerPro 216
QY 536 GCCTGTGACATCATCCACAGATGAATCATCCCGGGCTCACTGAGGCTCCTG---592
Db 217 LeuCysProAlaTyrr---ArgLysGluValValGlyLeuValAlaAspArgValCysLys 235
QY 593 ---TGCACTCAACAGCCCTTTGAGGTGGTGGAGAGGCTGCGCTCTCCTCAGCCACCTG 649
Db 236 LysCysProProGlnArgLeuAlaArgPheGluGluGluCysArgLysTyrrArgThrLeu 255
QY 650 GTGCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCCTCTACCCGCTGCTGAAGACCCC 709
Db 256 ValIleLysGlyValArgValPheAspValAlaValLeuAlaProLeuLeuArgAspPro 275
QY 710 TCCCTCAACCTCATATCGTCACCTGTGCGGAGACCCCGGCGCTGTTCGGTTCCCGA 769
Db 276 AlaLeuAspLeuLysValIleHisLeuValArgAspProArgAlaValAlaSerSerArg 295
QY 770 GAACGCACAAAGGAGATCTCATGATTGACAGTCGCAATTGTGATG-----814
Db 296 IleArgSerArgHisGlyLeuIleArgGluSerLeuGlnValValArgSerArgAspPro 315
QY 815 -----GGGACAGCATGAGCAAAACTCAAGAGGAG 844
Db 316 ArgAlaHisArgMetProPheLeuGluAlaAlaGlyHisLysLeuGlyAlaLysLysGlu 335
QY 845 -----GACCAACCTTACTAT-----GTGATGCAAGGTCTATCTGCCAAAGC 883
Db 336 GlyValGlyGlyProAlaAspTyrrHisAlaLeuGlyAlaMetGluValIleCysAsnSer 355
QY 884 CAGCTGGAGATCTACAAGACCATCCAGTCTTTCGCCAAGGCCCTGACAGAGCTTACCTG 943
Db 356 MetAlaLysThrLeuGlnThrAlaLeuGlnProProAspTrpLeuGlnGlyHisTyrrLeu 375
QY 944 CTTGTGCGCTATGAGGACCTGGCTCGAGCCCTGTGGCCAGACTTCCCGAATGTATGAA 1003
Db 376 ValValArgTyrrGluAspLeuValGlyGluProValLysThrLeuArgArgValTyrrAsp 395
```

QY 1004 TTCGTGGATGGAATCTTGGCCCATCTTCAGACCTGGTGATCAATCATCACCCGAGGC 1063
 Db PheValGlyLeuLeuValSerProGluMetGlnPheAlaLeuAsnMetThrSerGly 415
 QY 1064 AAGGCGATGGTGACACGCTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAG 1123
 Db SerGlySerSerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn 435
 QY 1124 GCTTGGCGCTGCTTGGCCCATCTTCGAAAGGTTTCTCGACTTCAGAAAGCCTGTGGCGAT 1183
 Db AlaTrpArgThrAlaLeuThrPheGlnGlnLeuValGluGluPheCysTyrGln 455
 QY 1184 GCCATGAATTTGGTGGCTACCGCCAGCTCAGATCTCAACAGACAGAAACCTGTG 1243
 Db ProMetAlaValLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer 475
 QY 1244 CTGGATCTTCG 1255
 Db LysThrLeuLeu 479
 RESULT 4
 A57397
 chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
 C:Accession: A57397
 R:Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi,
 J. Biol. Chem. 270, 18575-18580, 1995
 A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase
 A:Reference number: A57397; MUID:95355490; PMID:7629189
 A:Accession: A57397
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-458 <FUK>
 A:Cross-references: GB:D49915; MID:g971262; PIDN:BA08655.1; PID:g971263
 C:Superfamily: chondroitin 6-sulfotransferase
 C:Keywords: sulfotransferase
 Alignment Scores:
 Pred. No.: 1,03e-29 Length: 458
 Score: 500.50 Matches: 120
 Percent Similarity: 54.02% Conservatives: 75
 Best Local Similarity: 33.24% Mismatches: 133
 Query Match: 13.38% Indels: 33
 DB: 2 Gaps: 13
 US-09-645-078-1 (1-2043) x A57397 (1-458)
 QY 227 CTGCTATGAGGACACCCGAGCGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 286
 Db LeuGlyIleAlaAlaProGluProArgArgHisValLeuLeuMetAlaThrThrArgThr 123
 QY 287 GCGCTCTTCTTTGTTGGGCGAGCTTTTGGGCGAGCACCCGAGATGTTTCTACCTGATGGAG 346
 Db GlySerSerPheValGlyGluPhePheAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 143
 QY 347 CCGCGCTGGCAGTG-----TGGATGACCTTCAGACGAGCAGCCGCTGATGCTGCAC 400
 Db ProLeuTrpHisIleGluArgThrValThrPheGluProGlyGlyAlaAlaValGly 163
 QY 401 ATGGCTGTG-----CGGATCTGATACGGGCGCTCTTGTGCGCATGAGCGCTCTTT 454
 Db SerAlaLeuValTyrArgAspValLeuGlnLeuLeuLeuLeuLeuLeuLeuLeuLeu 183
 QY 455 GATGCTACATGAACCTGGTCCCGGAGACAG-----TCAGCGCTCTTTCAGTGGAG 508
 Db GluSerPheIleSerProAlaProGluGluHisLeuThrAlaAlaLeuPheArgArgGly 203
 QY 509 AACAGCGCGCGCTGTGTTCTGCACCTGCTGATCATCCACCAA---GATGAATC 565
 Db SerSerHisSerLeuCysGluGluProValCys-----ThrProSerLeuLysVal 221
 QY 566 ATCCCCCGGCTCACTGCGAG---CTCCTGTGCACTCAACAGCCCTTTGAGGTGGAG 622

Db 222 PheGluLysTyrHisCysLysAsnArgCysGlyProLeuAsnIleThrLeuAlaAla 241
 QY 623 AAGGCTCCCGCTCTCAGCCACGCTGTGTCTCAAGAGGTGCGCTTCTTCAACCTGCAG 682
 Db GluAlaCysArgArgLysGlnHisMetAlaLeuLysThrValArgIleArgGlnLeuGlu 261
 QY 683 TCCCTCTACCGCTGTCTGAAAGACCCCTCCCTCAACCTGCATATCGTCACCTGTCGG 742
 Db PheLeuGlnProLeuAlaGluAspProArgLeuAspLeuArgIleIleGlnLeuValArg 281
 QY 743 GACCCCGCGCGCTGCTCCGTTCCGAGAACGACAAAGGAGAGATCTCATGATTGACAGT 802
 Db AspProArgAlaValLeuValSerArg-----MetVal----- 292
 QY 803 CCATTTGATGGGCGACCATGAGCAAAAACCTCAAG----- 838
 Db ---AlaPheSerGlyLysTyrGluSerTrpLysLysTrpAlaAlaGluGlyGluAlaPro 311
 QY 839 ---AAGGAGGACCAACCTTACTATGTGATGAGCTCATCTGCCAAAGC---CAGCTGGAG 892
 Db LeuGlnGluAspGlu---ValGlnArgLeuArgGlyAsnCysGluSerIleArgLeuSer 330
 QY 893 ATCTAAGACCATCCATCCTTGGCCAAAGCCCTGACAGAACCTGCTGCTGTCGCGC 952
 Db AlaGluLeuGlyLeuArgGln---ProArgTrpLeuArgGlyArgTyrMetLeuValArg 349
 QY 953 TATGAGGACCTGGCTGAGCCCTGTCGCCAGACTTCCCGAATGTATGAATTCGTGGGA 1012
 Db TyrGluAspValAlaArgAlaProLeuArgLysAlaLeuGluMetTyrArgPheAlaGly 369
 QY 1013 TTGGAATTTCTGCCCATCTTTCAGACCTGGGTGTCATAACATCACCAGGAGGAGCATG 1072
 Db IleHisProThrProGlnValGluGluTrpIleArgAlaAsnThrGlnAlaPro---Gln 388
 QY 1073 GGTGACCGCTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGC 1132
 Db AspSerAsnGlyIleTyrSerThrGlnLysAsnSerSerGluGlnPheGluLysTrpArg 408
 QY 1133 TGTCTTTTCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGGGATGCCATGAAT 1192
 Db PheSerIleProPheLysLeuAlaGlnValValGlnAspAlaCysGluProAlaMetArg 428
 QY 1193 TTGCTGGCTACCGCCAGCTGATCTGACAGACAGAGAACCTGTTGCTGATCTT 1252
 Db LeuPheGlyTyrIleLeuAlaSerSerAlaGlnGluLeuThrAsnArgSerLeuSerLeu 448
 QY 1253 CTG 1255
 Db 449 Leu 449
 RESULT 5
 E95934
 probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sino
 C:Species: Sinorhizobium meliloti
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E95934
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95934
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CA49141.1; PID:gl5140626; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A;Reference number: A96039; MUID:21369234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMB21237
A;Genome: plasmid

[illegible]

QY 965 GCTCGAGCCCTGTGGCCAGACTCCCGAAGTATGAATTCGTGGGATTGGAATTCGTG 1024
 Db 225 AlaserAspProValAlaValMetGlnGlnIleGlyThrPheLeuGluLeuAsp---Leu 243
 QY 1025 CCCATCTTCAGACCTGGGTGCATACATCACCCGAGGCAAGGCATGGGT----- 1075
 Db 244 SerGlnValGlyThrSerLeuGluAsn-----GlyGluAlaMetGlyProGlyHis 260
 QY 1076 -----GACCACGCTTCCACACAAATGCGAGGATGCCCTTAATCTCTCCAGGCT 1126
 Db 261 GlnValAlaGlyAsnArgLeuArgMetAsnAlaSerIleAlaLeuAsnIlyAspGluThr 280
 QY 1127 TGGCGCTGGTCTTTGGCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGGCGATGCC 1186
 Db 281 TrpArgThrArgMetProAlaArgGlnGlnValSerPheGlnArgLeuGlyGlyTrpMet 300
 QY 1187 ATGAATTTGCTGGGCTAC 1204
 Db 301 LeuArgArgTyrGlyTyr 306

RESULT 6
 BHLF1 protein - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
 C:Accession: A03742
 R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85035713; PMID:6092825
 A:Accession: A03742
 A:Molecule type: DNA
 A:Residues: 1-660 <BAN>
 R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; F
 Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
 C:Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:	0.0058	Length:	660
Pred. No.:	147.00	Matches:	122
Score:	28.78	Conservative:	19
Percent Similarity:	28.78%	Mismatches:	175
Best Local Similarity:	24.90%	Indels:	174
Query Match:	3.97%	Gaps:	29
DB:	1		

	US-09-645-078-1	(1-2043) x QBE3	(1-560)	
QY	1211	CGTGGCGGTAGCCACGACAATTATGCATCCTGCCACAGGCTTTCTGAAGTCGAGAACCCT	1152	
Db	256	ArgSerGlyAlaAlaAlaGlnA ^G ThrHisA ^{Rg} -----ArgPro	268	
QY	1151	TTTTCATAGGCGCAAGACCAGCGCCAAAGCTTGGG-----AGACATTAA	1110	
Db	269	ProGlyCysProA ^T SerA ^I AarGasnProGlyCysProargThrTpA ^r ArgA ^r ArgSer	288	
QY	1109	GGGCATCCCCTGGCATTTTGCTGGAAGCGTGTCACCCATGCCCTCCCTGCCCTGGGTGARGT	1050	
Db	289	Gly-----AlaGlnA ^r G ^L yH ⁱ sProProProGlyAlaGly-----	300	
QY	1049	TATGCACCCAGGTCTGAAGATGGGGCAAGAAATTCATCCCACGAATTCATACATTCGGG	990	
Db	301	GlnArgProSerGlyProThrGlyL ^a r ^a rgProAlaAlaPro-----	314	
QY	989	AAGCTGGGCCACAGGGGCTCGAGCCAGGTCTCATACGGCACAGCAGGTAGCGTTCCCT	930	
Db	315	----GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaValPro	332	
QY	929	GCAGGG-----CCTTTGGCCAAGACTGGATGGTCTTTGTAGATCTCCAGCTGGCTTTGGC	876	

QY 35 TTCTGTCGTCTTCGGTACCCCAACAAATGG 6
DB 654 GlyCys-----ProArgThrTrp 659

RESULT 7
S21626
collagen alpha 1(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 13-Aug-1999
C:Accession: S57243, S16374, A23982; I49559; S39789; I48300; S21626
R:Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A:Reference number: S57243
A:Accession: S57243
A:Molecule type: mRNA
A:Residues: 1-1453 <LIS>
A:Cross-references: EMBL:U08020; NID:G470673; PIDN:AAA88912.1; PID:G470674
R:Wetstaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MET>
A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485
R:French, B.I.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A:Reference number: A23982; MUID:86137403; PMID:3841523
A:Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <PRE>
A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262
R:Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A:Reference number: I49559; MUID:83141374; PMID:6298597
A:Accession: I49559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264
R:Harbers, K.; Kuehn, M.; Dellus, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A:Reference number: I49557; MUID:84170331; PMID:6324198
A:Accession: I49557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <R2>
A:Cross-references: GB:X01688; NID:G192246; PIDN:AAA37330.1; PID:G553881
R:Fenton, S.P.; Laman, S.R.; Hamngan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8268229
A:Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80 'E', 82-105 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1
R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect
A:Reference number: I48300; MUID:94344105; PMID:8065328
A:Accession: I48300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80 'E', 82-105 'D', 107-147 <REF>
A:Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487
C:Genetics:
A:Gene: COL1A1
A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A>Title: Regulatory elements in the first intron contribute to transcriptional control c
A/Accession: A39943
A/Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>
A/Molecule type: DNA
A/Status: translated from GB/EMBL/DBJ
A/Residues: 1-34 <GB>
A/Cross-references: GB:J03559; NID:G180876; PIDN:AA52052.1; PID:G553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A/Reference number: 155237; MUID:85130970; PMID:2857713
A/Accession: 155237
A/Status: translation not shown; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-34 <CH2>
A/Cross-references: GB:J0627; NID:G180383; PIDN:AA51992.1; PID:G553226
R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A>Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
rome, type VII.
A/Reference number: A35233; MUID:90202908; PMID:2318955
A/Accession: A35233
A/Molecule type: protein
A/Residues: 33-52 <WR>
A/Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R:Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A>Title: A base substitution in the exon of a collagen gene causes alternative splicing
A/Reference number: S09400; MUID:89356643; PMID:2767050
A/Accession: S09400
A/Molecule type: mRNA
A/Residues: 156-183 <WB>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A/Reference number: A90567; MUID:71038625; PMID:5529814
A/Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
A/Accession: B90567
A/Molecule type: protein
A/Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A/Experimental source: skin
A/Note: evidence for 170-alanine
R:Baegge, B.; Norbom, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F
Eur. J. Biochem. 192, 153-159, 1990
A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A/Reference number: S11372; MUID:90382436; PMID:2169412
A/Accession: S11372
A/Molecule type: protein
A/Residues: 175-187, 274-287, 'P', 289 <BA>
A/Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantino, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
cooperative melting of intact type I collagen.
A/Reference number: 155342; MUID:92042092; PMID:1718984
A/Accession: 155342
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 258-268, 1347-1357 <DEA>
A/Cross-references: GB:S67495; NID:G239007; PIDN:AA20350.1; PID:G239008
A/Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A>Title: Comparative study of glycoproteins derived from selected vertebrate collagens.
A/Reference number: A92069; MUID:71001508; PMID:4319110
A/Accession: A92069
A/Molecule type: protein
A/Residues: 263-268 <MOR>
A/Experimental source: skin
A/Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A>Title: Segmental amplification of the entire helical and telopeptide regions of the cd

A/Reference number: S15989; MUID:90326017; PMID:2374517
A/Accession: S15989
A/Molecule type: mRNA
A/Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A/Reference number: 152905; MUID:93339042; PMID:8339541
A/Accession: 152905
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 342-352, 'C', 354-359 <W12>
A/Cross-references: GB:S64717; NID:G408195; PIDN:AA27677.1; PID:G408196
A/Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A>Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
A/Reference number: A90476; MUID:84080385; PMID:6689127
A/Accession: A90476
A/Molecule type: mRNA
A/Residues: 425-4250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A/Cross-references: GB:K01228; NID:G180391; PIDN:AA51995.1; PID:G180392
A/Note: sequence partially completed for missing nucleotides by A29439
R:Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A>Title: Multixon deletion in an osteogenesis imperfecta variant with increased type II
A/Reference number: A22161; MUID:85104934; PMID:2381843
A/Accession: A22161
A/Molecule type: DNA
A/Residues: 472-594, 'R', 596-607 <CH3>
A/Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AA51847.1; PID
A/Note: the authors translated the codon CGT for residue 595 as Pro
R:Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A>Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
A/Reference number: A35336; MUID:90252792; PMID:2339700
A/Accession: A35336
A/Molecule type: mRNA
A/Residues: 710-720, 'E', 722-737, 'B', 739-745 <WAL>
A/Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R:Porlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A>Title: Severe type III osteogenesis imperfecta due to glycine substitutions in the c
A/Reference number: 154365; MUID:95187161; PMID:7881420
A/Accession: 154365
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 746-766, 'S', 768-781 <FOR>
A/Cross-references: GB:47867; NID:G1009093; PIDN:AA59576.1; PID:G1009094
R:Chessier, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 269, 18218-18225, 1993
A>Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A/Reference number: A47426; MUID:93352646; PMID:8349697
A/Accession: A47426
A/Molecule type: mRNA
A/Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CH>
A/Cross-references: GB:S64596; NID:G407589; PIDN:AA27856.1; PID:G407590
A/Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBI:136445)
A/Note: does not represent an experimentally determined sequence but three different mut
A/Molecule type: mRNA
A/Residues: 1179-1464 <CH4>
A/Experimental source: normal dermal fibroblast culture
A/Accession: C47426
A/Molecule type: mRNA
A/Residues: 1179-1276, 'H', 1278-1464 <CH5>
A/Experimental source: fetal cell 86-237
A/Accession: D47426
A/Molecule type: mRNA
A/Residues: 1179-1336, 1339-1464 <CH6>
A/Experimental source: fetal cell 86-146
A/Accession: E47426

333 ProProGlyProThrGlyProAlaGly-----ProProGlyPheProGly-Al 348
QY 641 TGTAGGACGGCAGCGCCCTTCCACCACCTCAAAAGCGTGTTGAATGCACAGGAGCCTTCG 582
Db 348 aValGlyAla-----LysGlyGluAlaGlyProGlnGlyProAr 361
QY 581 AGTGAGCCCGGGGATGATTTCATCTTGTGGGATGATGTCACAGGCAGGTGCAGAACACA 522
Db 361 gCySerGluGlyPro-----GlnGlyValArgGlyGluProGlyProPr 376
QY 521 GGGCCCCGGCTGTCTCCCACCTGAAGAGGCTGGACTG-----TCTCGGGGACACAGGTT 468
Db 376 odLyProAla-----GlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGl 392
QY 467 CCATGTAGGCATCAAAGACGCTCATGTCGCAAGAAGACGCCCGGTATCAGATCCCCGCA 408
Db 392 yclnProGlyAlaLysGlyAlaasnGlyAla----- 402
QY 407 CAGCCATGTGCAGCATCCAGC-----ProdglyleAlaGlyAlaProGlyPheProGlyAlaAargGlyPr 417
Db 403 -----ProdglyleAlaGlyAlaProGlyPheProGlyAlaAargGlyPr 417
QY 371 AGGTCATCACACGTGCCAGCGGGCTTCATCAGGTAGTAACATCTGGGTGCTGCCCAA 312
Db 417 oSerGlyProGlnGlyProGlyGly--ProProGlyProLysGly---AsnSerGlyGlu 435
QY 311 AAAGTGCGCCACAAAAGAGACGACAGAGCGCCGAGGAACACAGACACGTCGATGC 252
Db 436 ProGlyAlaProGlySerLysGlyAspThrGlyAlaLysGlyGluProGlyProValGly 455
QY 251 GCTCGGGCTGTGCCTTCATAGACAGGAGAGCTGATGTGTGGCTGTACATGTGGGAAGATA 192
Db 456 ValGlnGlyProProGlyProAlaGlyGlu-----GluGlyLysArgGlyAlaAargGly 473
QY 191 GAGCCAAGATGGCCATCTGGGAAACACAGAAACAGCAGGAGGACTTCATTTTTTAGSCAGTA 132
Db 474 GluPro----GlyProThrGlyLeuProGlyProProGlyGlu----- 486
QY 131 GCATGTGCTCAAGTGGAAAGACCTTGTGGCTTGGGCTTCCTCGCTGCT 84
Db 487 -----ArgGlyGlyProGlySerArgGlyPheProGlyAla 498

RESULT 9
TSHP2
thrombospondin 2 precursor - human
C;Species: Homo sapiens (man)
C;Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C;Accession: A47379; A42173
R;LaBell, T.L.; Byers, P.H.
Genomics 17, 225-229, 1993
A;Title: Sequence and characterization of the complete human thrombospondin 2 cDNA sequence, chromosome location, and expression pattern of the gene.
A;Reference number: A47379; MUID:94010892; PMID:8406456
A;Molecule type: mRNA
A;Residues: 1-1172 <LAAB>
A;Cross-references: GB:I42350; NID:g307505; PIDN:AAAO3703.1; PID:g307506
R;LaBell, T.L.; Milewicz, D.J.; Distcheke, C.M.; Byers, P.H.
Genomics 12, 421-429, 1992
A;Title: Thrombospondin II:
A;Reference number: A42173; MUID:92217961; PMID:1559694
A;Accession: A42173
A;Molecule type: mRNA
A;Residues: 560-1172 <LA2>
A;Cross-references: GB:M61339
A;Experimental source: fibroblast
A;Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:P.95096)
C;Genetics:
A;Gene: GDB:THB82; TSP2
A;Cross-references: GDB:128789; OMIM:188061
A;Map position: 6q27-q27
C;Complex: homotrimer, disulfide linked
C;Function:
A;Description: participates in cell migration and adhesion, and in platelet aggregation.

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vcd
 F:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; trimmed
 F:18/Domain: signal sequence #status predicted <SIG>
 F:19-1172/Product: thrombospondin 2 #status predicted <WAT>
 F:319-377/Domain: von Willebrand factor type C repeat homology <WVC>
 F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
 F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
 F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
 F:553-588/Domain: EGF homology <EGF1>
 F:552-691/Domain: EGF homology <EGF>
 F:928-930/Region: cell attachment (R-G-D) motif
 F:351,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:167-226/Disulfide bonds: #status predicted
 F:266,270/Disulfide bonds: interchain #status predicted
 F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Alignment Scores:

Pred. No.: 0.0121 Length: 1172
 Score: 143.50 Matches: 131
 Percent Similarity: 29.09% Conservative: 61
 Best Local Similarity: 19.85% Mismatches: 229
 Query Match: 3.83% Indels: 239
 DB: 1 Gaps: 31

US-09-645-078-1 (1-2043) x TSHUP2 (1-1172)

QY 199 CCACATGACAGCCACACATCAGCTCCCTGCTTATGAAGGCACACCCGAG----- 250
 Db 248 ProHisValThrThrGluThrValGlyPro-SerSerGluArgProGluValCysG1 267
 QY 251 -----CGCATGCACGTGCTGCT 267
 Db 267 uArgSerCysGluGluLeuGlyAsnMetValGlnGluLeuSerGlyLeuHisValLeuVa 287
 QY 268 T-----CTGCTCTCCGCGCTCTGGCTCTCTTTGTGGGGCAGCT 309
 Db 287 lAsnGlnLeuSerGluAsnLeuLysArgValSerAsnAspAsnGlnPheLeuTrpGluLe 307
 QY 310 TTTTGGGCAGCACCCAGATGTTTCTACCTGATGGAGCCCGCTGGCAGCTGTGGATGAC 369
 Db 307 uileGlyGlyProProLysThr---ArgAsnMetSerAlaCysTrpGlnAspLysArgph 326
 QY 370 CTTCAACGACAGACACCCCTGGATGTCACATGGCTGTGCGGGATCTGATACGGCCGT 429
 Db 326 ePheAlaGluAsnGluThrTrp-----ValValAspSerCys 338
 QY 430 CTTCTTGTGGACATGAGCGTCTTTGATGCTATACGACCTGGTCCCGGAGACAGTC 489
 Db 338 sThrThrCysThrCysLysLysPhe----- 346
 QY 490 CAGCCTCTTTTCAGTGGGAGAACACCGCGCCCTGTGT-----TCTCACCTGCTG 540
 Db 347 -----LysThrLeCysHisGlnIleThrCysPro-ProA 358
 QY 541 TGACATCATCCACAGATGAATCATCCCGGGCTCATCTGACGGCTCCTGT----- 593
 Db 358 lathrCys-----AlaSerProSerPheValGluGlyCysCysProS 373
 QY 594 -----GCAGTCAACAGCCCTTTGAGGTGGTGGAGAGGCGCTGCGCTCCTACAG 642
 Db 373 erCysLeuHisSerValAspGlyGluGluGlyTrp-----SerProTrpA 388
 QY 643 CCACGTGG-----TGCTCAAGGAGGTGCGCTTCTTCACTGACGTCCCTCTACCCGCT 696
 Db 388 laGluTrpThrGlnCysSerValThrCysGlySerGlyThrGlnGlnArgGlyArgSerC 408
 QY 697 GCTGAAGACCCCTCCCTCAACCTGCATA-----TCGTGACAC- 734
 Db 408 ysaspValThr---SerAsnThrCysLeuGlyProSerIleGlnThrArgAlaCysSerL 427
 QY 735 -----TCGT 738
 Db 427 euserLysCysAspThrArgIleArgGlnAspGlyGlyTrpSerHisTrpSerProTrps 447

QY 739 CCGGAGCCCCCGCGCTGTTCGCTTCCGAGAACGCCAACAGGGAGATCTCATGATTGA 798
 Db 447 exSerCysSerValThrCysGlyValGlyAsnIleThrArgIleArgLeuCysAsnSerP 467
 QY 799 CAGTCGATTTGATGGGCGAGCATGACAAAGAACTCAAGAGGAGGAGCAACCCCTACTA 858
 Db 467 roValProGlnMetGlyGlyLysAsnCysLysGlySerGlyArgGluThrLysAlaCysG 487
 QY 859 TGTGATGTCAGGTTCATCTGCCAAAGCCAGCTGGAGATCTACAGACCATCAGTCTCTGCC 918
 Db 487 lnglyAlaProCysProIleAspGlyArgTrpSerProTrpSerProTrpSerAlaCysT 507
 QY 919 CAAGGCCCTGC-----AGGAAGCTACCTGCTTGTGCGCTATGAGACCTGGC 966
 Db 507 hrValThrCysAlaGlyGlyIleArgGluArgThrArgValCys----- 521
 QY 967 TCGAGCCCTGTGGCCCGAGACTTCCCGAA-----TGATGAAAT 1005
 Db 522 -----AsnSerProGluProGlnThrGlyGlyLysAlaCysValGlyVa 536
 QY 1006 CG-----TGGATTGGAAATCTTGGCCCATCTTCAGACCTGGGTGCA 1047
 Db 536 spValGlnGluArgGlnMetCysAsnLysArgSerCysProVal-----AspGlyCysL 554
 QY 1048 TAACATCACCCGAGGCAAGGCGATGGGTGACACGCTTTCCACACAAATCCAGGGATGC 1107
 Db 554 euserAsnPro-----CysPheProGlyAlaG 563
 QY 1108 CCTTAATGCTCCCGAGCTTGGCGCT-----GGTCTTTCCTCATGAAAGGTTTC 1158
 Db 563 lncysSerSerPheProAspGlySerTrpSerCysGlyPheCysProVal---GlyPheL 582
 QY 1159 TC-----GACTTCAGAAAGCCTGTGGCGATGCCATGATGATTTGCT 1197
 Db 582 euGlyAsnGlyThrHisCysGluAspLeuAspGluCysAlaLeuValProAspIleCysP 602
 QY 1198 GGGTACCGCCACGCTCAGATCTGAACAGAACACAGAGAACTGTGTGATCTTCTGTC 1257
 Db 602 he-----SerThrSerLysValProArgCysValAsnThrGlnProGlyPheHisCysL 620
 QY 1258 TACCT-----GGACTGTCCCTGAGCA 1278
 Db 620 euproCysProProArgTrpArgGlyAsnGlnProValGlyValGlyLeuGluAlaLaL 640
 QY 1279 AATCCACTAGAGGTTGAGAAAGCTTTGTGCGCACCTGGTGTGCGCTCAGTCACTTTC 1338
 Db 640 ysThrGluLysGlnValCysGluProGluAsnPro----- 651
 QY 1339 TCTGATGCTTCTGAGCCTTGCCTACATCTCTGAGCCTTAACATACATGCTCTGGGTATC 1398
 Db 652 -----CysLysAspLysThrHisAsnCysHisLysH 662
 QY 1399 ACACGTAGTGTGAGTTGTCCACACGCTGCTCAAGCAGAGAGACTTTTGTGTCCATGCTT 1458
 Db 662 lAlaGluCysIleThrLeuGlyHisPhe-SerAspProMet-TyrLysCysGluCys-- 680
 QY 1459 GTGCTAGAAAACAGACTGGGGAACCTTATGTGAGCAGCACATCCACCCAGTGAACAGG 1518
 Db 681 -----GlnThrG1 683
 QY 1519 GTATTGCTTCTTCTTCTTCTT-----GATCTTCTGTCTGGCAGACTT 1563
 Db 683 yTyAlaGlyAspGlyLeuIleCysGlyGluAspSerAspLeuAspGlyTrpProAsnLe 703
 QY 1564 CAGAGACTTTGGGCTGGAGGCTTATTAGCAGCAGACATATCAGTGGATGATCA 1623
 Db 703 uAsnLeuValCysAlaThrAsnAlaThrTyHis-----CysI1 716
 QY 1624 TAAACCTCCCTCTCCACATCTTCCCAATGGGAATGGATCTTTCCACAAAGAGCTACC 1683
 Db 716 elysAspAsnCysProHisLeuProAsnSerGlyGlnGluAspPheAspLysAspGlyI1 736

QY 1684 AGCATTTTCCACAGAG----- 1699
Db 736 eGlyAspAlaCysAspAspAspAspAspGlyValThrAspGluLysAspAsnC 756
QY 1700 -ATCGGAATCTGAGCCCTGGAGTCCCAATCGGATTCAGGAAGAGTGGGACAG 1758
Db 756 sGlnLeuLeuPheAsnProArgGlnAlaAspTyrAsp---LysAspGluValGlyAspAr 775
QY 1759 GTTGGATGCTACTTATGAGCTTGACCATACAGCTATCGGTAATCAGAAATA 1810
Db 775 sCysAspAsnCys-----ProTyrValHisAsnProAlaGlnIle 788
RESULT 10
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 13-Aug-1999
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:Cross-references: EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, 'G', 868-1464 <TOA>
A:Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57983; NID:950476; PIDN:CAA1048.1; PID:950477
C:Genetics:
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 295/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3
S:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-154/Domain: propeptide #status predicted <PRO>
F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>
F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Alignment Scores:
Pred. No.: 0.0213 Length: 1464
Score: 140.50 Matches: 97
Percent Similarity: 32.57% Conservative: 17
Best Local Similarity: 27.71% Mismatches: 130
Query Match: 3.80% Indels: 106
DB: 2 Gaps: 20
US-09-645-078-1 (1-2043) x S59856 (1-1464)
QY 1124 CTGGAGACATTAAGGGCATCTCGGCAATTTGTGTGGAAAGCTGGTCAACCATGCCCT 1065
Db 814 ProGlyGlnAsn---GlyGluPro---GlyAlaLysGlyGluArgGlyAlaPro----- 829
QY 1064 TGCCCTGGGTGATGTTATCCACCCAGTCTGAAGTCTGAAGTGGGGCAGAATTCATCCACCA 1005
Db 830 -----GlyGluLysGlyGluGlyPro-----ProGlyProAla 841
QY 1004 ATTCAATACATTCGGGAAGTCTGGGCCACAGGGCTCGAGCCAGGCTCTCATAGCCACAA 945
DB: 11 Indels: 106 Gaps: 20

Db 842 GlyProThr---GlySerSerGlyProAlaGly---ProProGlyProGlnGlyValLys 859
QY 944 GCAGGTAGCGTTCTCTGAGGGCCCTTGGCAGGACTGGATGCTTCTTAGATCTCCAGCT 885
Db 860 GlyGluArgGlySerProGlyProGlyThrAlaGlyPhe-----ProGly 875
QY 884 GCGTTTGGC-----AGATGACCTGCATCACATAGTAGGGTTGGTCTCTCTTCTTGAGTT 831
Db 876 GlyArgGlyLeuProGlyProProGlyAsnAsnGlyAsnProGlyProProGly- 893
QY 830 TTTGCTCATGCTGCCCATCAATGCGACTGTCTCAATCATGAGATCTCCCTTTGCGGTT 771
Db 894 ---ProSerGlyAlaProGlyLys-----AspGlyProProGlyPro 906
QY 770 CTCGGGAACGGAACACGCGCGGGGTCCTCGGA----- 738
Db 907 AlaGlyAsnSerGlySerProGlyAsnProGlyLeuAlaGlyProLysGlyAspAlaGly 926
QY 737 -----CCAGGTGCACGATATGAGGTTGAGGGAGGGGT---CTT 702
Db 927 GlnProGlyGluLysGlyProProGlyAlaGlnGlyProProGlySerProGlyProLeu 946
QY 701 TCAGCAGCGGGT-----AGAGGACTCGAGTTCGAGGAGGCGCCTCTTGAAGAGCGCCTCTTGA 654
Db 947 GlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGly----- 959
QY 653 GCACCACTGCTGTAGGAGCGGCGGCTTCTCCACCATCAAGAGGCTGTTCAGTGC 594
Db 960 ---ProProGlyMetProGlyProArgGlySerProGlyProGln-----GlyIle 975
QY 593 ACAGGAGCTCGAGTGCAGCCCGGGGATGATTCATCTTGTGGATGATGTCACAGGCGAG 534
Db 976 LysGlyGluSerGlyLysProGlyAlaSerGlyHis-----Asn-GlyGly 990
QY 533 GTGCAGAACACAGGCGCGGCTTCTCCACTGAAGAGGCTGAGTGTCTCCGGGAGC 474
Db 990 uArgGlyProProGlyProGlnGlyLeuProGlyGln-----ProGlyTh 1005
QY 473 CAGGTTCCATGTAGGATCAAGAGCGCTCATGTCACCAAGAGAGCGCCGCTATCAGAT 414
Db 1005 rAlaGlyGluProGly---ArgAspGlyAsnProGlySerAspGlyGlnPro----- 1021
QY 413 CCCGCACACCCATGTCAGCATCCAGGCGGTCTCTGTTGAAGGTCATCCACAGTGC 354
Db 1022 -----GlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluAsnGlySerPr 1039
QY 353 AGCGGGCTCCATCAGGTAGAAAACATCTGGTGTGCTGCCAAAAGCTGCCCAACAAAG 294
Db 1039 oGly--AlaProGlyAlaProGlyHisProGlyProProGlyProValGlyProSerGly 1058
QY 293 AACAGCCAGAGCGCCAGGAGAGCAGAACCCAGCAGCGTGCATGCGCTCGGCTGTGCTTCA 234
Db 1059 LysSerGlyAspArgGlyGluThrGlyProAla----- 1069
QY 233 TAGACAGGGAGCTGATGTTGTGCTGTACATGTGGAGAATAGACCAAGATGGCCATCT 174
Db 1070 -----GlyProSer 1072
QY 173 GGGAAACCAAGAACACAGAGAGCT 150
Db 1073 GlyAlaProGlyProAlaGlyAla 1080
RESULT 11
JE0291
FB19 protein - human
C:Species: Homo sapiens (man)
C:Date: 05-Dec-1998 #sequence revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: JE0291
R:Totaro, A.; Grifa, A.; Carella, M.; Rommens, J.M.; Valentino, M.A.; Roetto, A.; Zelant
Biochem. Biophys. Res. Commun. 250, 555-557, 1998
A:Title: Cloning of a new gene (FB19) within HLA class I region.
A:Reference number: JE0291; MUID:99003493; PMID:9784381
A:Accession: JE0291

A:Reference number: I38867; MUID:95150028; PMID:7847372
 A:Accession: I38867
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TII>
 A:Cross-references: EMBL:U15195; NID:9557053; PIDN:AAB60370.1; PID:9557054
 R:Ramirez, P.
 submitted to the EMBL Data Library. December 1988
 A:Reference number: S04892
 A:Accession: S04892
 A:Molecule type: mRNA
 A:Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <RAM>
 A:Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050
 R:Viikula, M.; Peltonen, L.
 FEBS Lett. 250, 171-174, 1989
 A>Title: Structural analyses of the polymorphic area in type II collagen gene.
 A:Reference number: S05000; MUID:89325561; PMID:2753125
 A:Accession: S05000
 A:Molecule type: DNA
 A:Residues: 630-640, 'A', 642-785 <VIK2>
 A:Cross-references: EMBL:X16158; NID:929951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
 PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
 R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D.
 J. Biol. Chem. 267, 22522-22526, 1992
 A>Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
 A:Reference number: A44309; MUID:93054548; PMID:1429602
 A:Accession: A44309
 A>Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA; mRNA
 A:Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',
 A:Cross-references: GB:100977; NID:g180812; PIDN:AB23914.1; PID:g258774
 A>Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence we
 A>Note: this translation is not annotated and this publication is not cited in GenBank
 A>Note: mutant sequence associated with perinatal lethal hypochondrogenesis
 R:Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
 A>Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
 A:Reference number: S16502; MUID:90251662; PMID:2339128
 A:Accession: S16502
 A:Molecule type: DNA
 A:Residues: 1164-1184, 'GPSKGKGGANGIPGI', 1185-1199 <TII2>
 A:Cross-references: EMBL:M37126; NID:9180808; PIDN:AAA52037.1; PID:g180809
 A>Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
 R:Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
 Proc. Natl. Acad. Sci. U.S.A. 92, 2555-2559, 1995
 A>Title: Identification and characterization of the human type II collagen gene (COL2A1)
 A:Reference number: A02858; MUID:85190534; PMID:3857598
 A:Accession: A02858
 A:Molecule type: DNA
 A:Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487 <CHE>
 A:Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
 R:Elima, K.; Vuorio, T.; Vuorio, E.
 Nucleic Acids Res. 15, 9499-9504, 1987
 A>Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
 A:Reference number: A27280; MUID:88067771; PMID:2825137
 A:Accession: A27280
 A:Molecule type: DNA; mRNA
 A:Residues: 1175-1487 <ELI>
 A:Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:g30097
 A:Experimental source: fetal epiphyseal cartilage
 R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
 Biochem. J. 237, 923-925, 1986
 A>Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
 A:Reference number: A57033; MUID:87099927; PMID:3800925
 A:Accession: A57033
 A:Molecule type: protein
 A:Residues: 'XS', 1244-1246, 'N', 1248, 'X', 1250-1255; 1295-1305; 1395-1408 <VAN>
 A>Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal prope
 R:Strom, C.M.; Upholt, W.B.
 Nucleic Acids Res. 12, 1025-1038, 1984
 A>Title: Isolation and characterization of genomic clones corresponding to the human typ
 A:Reference number: A21733; MUID:84118798; PMID:6320112
 A:Accession: A21733

A:Molecule type: DNA
 A:Residues: 1245-1295 <STR1>
 A:Cross-references: EMBL:X00339; EMBL:X00298; NID:9394699; PIDN:CAA25092.1; PID:94378975
 A:Accession: B21733
 A:Molecule type: DNA
 A:Residues: 894-909, 'PE', <STR2>
 A:Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
 R:Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
 Biochemistry 24, 6343-6348, 1985
 A>Title: Isolation and partial characterization of genomic clones coding for a human pro
 Gene.
 A:Reference number: A24561; MUID:86104139; PMID:3002437
 A:Accession: A24561
 A:Molecule type: DNA
 A:Residues: 1296-1358 <NUN2>
 A:Cross-references: GB:M12048; NID:g180017
 A>Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
 A>Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
 R:Sangiorzi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
 Nucleic Acids Res. 13, 2207-2225, 1985
 A>Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
 A:Reference number: I37249; MUID:85215609; PMID:2987845
 A:Accession: S59491
 A:Molecule type: DNA
 A:Residues: 7-28; 'R', 99-114, 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1
 A:Accession: I84453
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 7-28 <SAN2>
 A:Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
 A>Note: the GenBank PID is based on an incorrect reading frame
 A:Accession: I37250
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 541-560 <SAN3>
 A:Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
 A:Accession: I37251
 Alignment Scores:
 Pred. No.: 0.0303 Length: 1487
 Score: 138.50 Matches: 87
 Percent Similarity: 35.47% Conservative: 18
 Best Local Similarity: 29.39% Mismatches: 119
 Query Match: 3.74% Indels: 72
 DB: 1 Gaps: 17
 US-09-645-078-1 (1-2043) x CGHU6C (1-1487)
 QY 1142 GCAAGACACGCGCC---AAGCGTGGGAGACATTAGGGCATCCCTCGCATTTGTGTGA 1086
 DB 418 AlaLysGlySerAlaGlyAlaProGly-----IleAlaGly 429
 QY 1085 AAGCGTGTACCCATGCCCTTGCCTGGGTGATGTTATGACCCAGGTCTGAAGATGGG 1026
 DB 430 AlaProGlyPheProGlyPro-----ArgGlyProProAspProGlnGly 444
 QY 1025 GCAAGAATTCCAATC-----CCACGAATTATACATTC--- 993
 DB 445 AlaThrGlyProLeuGlyProLysGlyGlnThrGlyLysProGlyIleAlaGlyPheLys 464
 QY 992 GGAAGTCTGGCCACAGGGCTGAGCCAGGTCTCTATAGCCACAGCAGGATGGGTT 933
 DB 465 GlyGluGlnGlyProLysGly---GluProGlyProAlaGlyProGlnGlyAlaProGly 483
 QY 932 CCGCAGGGCCCTTGGGCAAGGACTCGATGCTTGTAGATCTCCAGGTGGCTTTGGCAGA 873
 DB 484 ProAlaGlyGluGluGlyLysArgGly---AlaArgGlyGluProGlyValGly--- 501
 QY 872 TGACCTGCATCACATAGTAGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 825
 DB 502 -----ProIleGlyProGlyGluArgGlyAlaProGlyAsnArg 515
 QY 824 -----CATGCTGCCCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 798

```
Db 516 GlyPheProGlyGlnAspGlyLeuAlaGlyProLysGlyAlaProGlyGluArgGlyPro 535
QY 797 CAATCATGATCTCCCTTGTGCGTCTCGGAACGAAACACGCGCCGGGGTCCCGGA 738
Db 536 SerGlyLeuAlaGlyProLysGlyAlaAsnGlyAspProGlyArgProGlyGluProGly 555
QY 737 ---CCAGGTGCACGATATGCGAGGTGAGGAGGGTCTTTCCAGACGCGGGTAGAGGACT 681
Db 556 LeuProGlyAlaArgGlyLeuThr---GlyArgProGlyAspAlaGlyProGlnGlyLys 574
QY 680 GCAGTTGAAGAGCGACCTCTTGTGAGCACACCGTGTGTAGAGCGGCGGCTTCT 621
Db 575 ValGlyProSerGlyAlaProGly-----GluAspGlyArgProGly 588
QY 620 CCACCA---CCTCAAGGCGTGTGACTGCACAGGAGCTCAGTGCAGCGCGGGATGA 564
Db 589 ProProGlyProGln-----GlyAlaArgGlyGlnProGly----- 600
QY 563 TTTCATCTTGGGGATGATGCACAGGAGGTGCAGACACAGGCGCGGCTGTCTCC 504
Db 601 -----ValMetGlyPheProGlyProLysGlyAlaAsnGlyGluProGlyLysAlaGly 618
QY 503 ACTGAAGAGGCTGACTGTCTCGGGACACAGGTTCCATGAGCATCAAGAGGCTCA 444
Db 619 Glu-LysGlyLeuProGlyAlaProGlyLeuArgGlyLeuProGly---LysAspGlyG 637
QY 443 TGTGCACAAAGAGCGCGCGTATCAGATC-----CGCACACGCCATGTGCA 396
Db 637 uThrglyAlaGlyProProGlyProAlaGlyProAlaGlyGluArgGlyGluGlnG 657
QY 395 GCATCAGCGGTGCTGCTGTGAGTCAAGTCCACACGTCGACGCG 350
Db 657 yAlaProGlyProSerGlyPheGlnGlyLeuProGlyProGly 672

RESULT 14
F75420
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75420
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036996; PMID:10567266
A:Accession: F75420
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-319 <WHI>
A:Cross-references: GB:AE001971; GB:AE000513; NID:G6458972; PIDN:AAF10810.1; PID:G645898
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1229
A:Map position: 1

Alignment Scores:
Pred. No.: 0.0238 Length: 319
Score: 138.00 Matches: 79
Percent Similarity: 30.82% Conservative: 15
Best Local Similarity: 25.90% Mismatches: 87
Query Match: 3.69% Indels: 124
DB: 2 Gaps: 23

US-09-645-078-1 (1-2043) x F75420 (1-319)

QY 219 TCAGTCCCTGCTCATGAGGACAGCCGAGCGATGACGTGCTGG----- 266
Db 77 SerAlaProCys-----ProSerAlaAlaGlyCysTrpAsnSerSerAla 91
QY 267 -----TTCTGTCTCTGCGCTGCTGGCTCTCTTTTCTGGGCG 305
```

```
Db 92 ArgValArgAspArgArgSerCysSerProSerArgArgCys-----TrpGly 108
QY 306 AGCTTTTGGGAGACACCCAGATGTTTCTACTGATGGAGCCGCTGGCAGGTGGA 365
Db 109 -----TrpProSerLysGly-----CysArg 115
QY 366 TGACCTTCAAGCAGACACCGCTGGA-----TGC 395
Db 116 ---ProGlyLysArgGlyGlyAlaGlyTTPProArgArgSerLysThrTrpCys 134
QY 396 TGCACATGCTGTGGGGATCTGATACGGCGCTTCTTGTGGCAGATGACGCTTTG 455
Db 135 GlyGlyTrpArgAlaAlaSerThrSerGlySerSerCysAlaAlaArgLysArg 154
QY 456 ATGCTTACATGAACCTGTCGCCGAGACACAGCTCCAGCTCTTTCAGTGGGAGAACAGCC 515
Db 155 ArgArgHisTrp-----GlySerSerGlyProSer---SerGlyCysSerAla 169
QY 516 GGGCCCTGT---GTTCTGCACCTGCTGTGACATCATCCCAAGATGAATCATCCCCC 572
Db 170 AsnThrThrArgAlaArgHisAlaProAla-----ArgSerSerPro 183
QY 573 GGGCTCCTCAGGCTCCTGTGCAGTCAACAGCCCTTTGAGG-----TGGTGGAGAAGG 626
Db 184 LeuAlaSerThrGly-----HisSerProAlaAlaSerGlyTrpGlyArg--- 198
QY 627 CTGCGCTCTCTACACCCACGTGTGC-----TCAAGGAGGTGCGCTTTCACACCTGC 680
Db 199 -----SerProSerArgAsnTrpCysGlyAlaAlaArgProSerValSerSerGlyArg 216
QY 681 AGTCCCTCTACCGCTGCTGAAAGACC-----CCTCCCTCA 716
Db 217 AlaThrSerThrArgCysTrpProAlaAlaSerAlaAsnSerValProLysProSer 236
QY 717 ACCTGCATATCG-----TGACCTGTGTCGGGACCCCC 749
Db 237 SerProArgSerValArgAlaGlyThrArgTrpGlyCysCysThrTrp-----ThrPro 254
QY 750 GGGCGCTGTCGTTCCCGAGAACGACACAAAGGAGATCTCATGATTGACATGCGCATG 809
Db 255 GlyAlaProAlaAlaProArg----- 261
QY 810 TGATGGGCGAGCATGAGCAAAAACCTCAGAGAGGAGGACCAACCTTACTATGTGATGC--- 866
Db 262 -----AlaThrThrThrArgAsnSer-----CysSer 270
QY 867 AGTCTATCTGCCAAGCCAGCTGGAGA----- 893
Db 271 HisSerProSerArgArgValTrpArgLeuThrThrProGlySerThrProSerArgPro 290
QY 894 ---TCTACAGACCATCCAGTCTTGC-----CCAAGGCGCTGCAGGAGAACCTACCTGC 944
Db 291 AlaSerAlaArgProProArgArgCysAlaArgProAlaProCysTrpArgAlaAlaCys 310
QY 945 TTGTGCGCTATGAGG 959
Db 311 ThrCysProThrArg 315

RESULT 15
CGR15
collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C:Accession: A90559; A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209; A91
R:Bornstein, P.
Biochemistry 8, 63-71, 1969
A:Title: Comparative sequence studies of rat skin and tendon collagen. II. The absence c
A:Reference number: A90559; MUID:69155173; PMID:5777344
A:Contents: CNB:0 and CNB:1
A:Accession: A90559
A:Molecule type: protein
A:Residues: 1-19 <BO1>
A:Experimental source: tendon
```

A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) in
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FBS Lett. 26, 61-65, 1972
A;Title: Non-helical regions in rat collagen alpha1-chain.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Accession: CNB+6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <SW2>
A;Experimental source: skin
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (in
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNB+ peptides in the alpha 1(I) chain of rat skin colla-
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl-
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status
F;9/Modified site: allylsine (lys) #status experimental
F;103/424,547/Binding site: carbohydrate (lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (lys) (partial) #status experimental
F;424,547/Modified site: 5-hydroxylysine (lys) (partial) #status experimental

Alignment Scores:
Pred. No.: 0.0279 Length: 671
Score: 138.00 Matches: 101
Percent Similarity: 28.50% Conservative: 19
Best Local Similarity: 23.99% Mismatches: 137
Query Match: 3.73% Indels: 165
DB: 1 Gaps: 23

US-09-645-078-1 (1-2043) x CGRT1S (1-671)

Qy	1144	GGGCAAGACCAGCGGCAAGCTGGAGACATTAAGGGCATCCCTGGCAT-----	1095
Db	80	GlyGlnArgGlyProGlyProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu	99
Qy	1094	-----TTGTGTGGAAGCGTGGTCAACCATGC	1068
Db	100	ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAsnThrGly	119
Qy	1067	CCITGGCTCGGTGATGTTATGCACCCAGGTCTGAAGTGGGCA-----AGA	1020
Db	120	ProAla-----GlyProLysGlyGluProGlySerProGlyGluAsnGly	134
Qy	1019	ATTCCATCCCGAATTCATACATTCGGGAAGTCTGGCCACAGGGGCTC-----	969
Db	135	AlaPro-----GlyGlnMetGlyProArgGlyLeuProGlyGlu	147
Qy	968	-----GAGCCAGTCTC-----CATAGCCACACAAGCA	942
Db	148	ArgGlyArgProGlyProProGlySerAlaGlyAlaArgGlyAspAspGlyAlaValGly	157
Qy	941	GGTAGCCTTCCTCGAGGGCCCTCGGCAAGACTGGATGCTTTGTAGATCTCCAGTGGC	882
Db	168	AlaAlaGlyProProGlyProThrGlyProThrGly-----ProProGly	182
Qy	881	TTTGGCAGATGACCTGCATCACATAGTAGGGTTGGTCTC-----	843
Db	183	PheProGlyAlaAlaGlyAlaLysGlyGluAlaGlyProGlnGlyAlaArgGlySerGlu	202
Qy	842	---CCTTCTTGAGTTTTTGTCTCATGCTGCCCCCATCAATCGCACTGTCAATCATGAGAT	786
Db	203	GlyProGlnGlyValArgGlyGluProGlyProProGly-----	215
Qy	785	CTCCCTTTGTGCGTCTTCGGGAACGGAACACGGCCGGGGGTCCCGCA---CCAGGTGCA	729
Db	216	---ProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAla	234
Qy	728	CGATATGCAGGTTGAGGAGAGGGGCTTTTCACAGCG-----GGTAGAGGG	684

```

Db 235 LysGlyAlaAsnGlyAlaProGlyPheProGlyAlaArgGly 254
    : : : | | | : : : : | | | | |
QY 693 ACTGAGGTTGAAGAGCGACCTCTGAGCAGCAGTGGCTGT----- 639
Db 255 ProSerGly-----ProGlnGlyProSerGlyAlaProGlyProLysGly 269
    : : : | | | : : : | | | | |
QY 638 AGGAGCGGCGAGCCTTCTCCACCACCTCAAGGGCTGTTCACATGCACAGGAGCCTGCAGT 579
    : : : | | | : : : | | | | |
Db 270 AsnSerGlyGluProGlyAlaProGlyAsnLysGly-----AspThrGlyAlaLysGly 287
    : : : | | | : : : | | | | |
QY 578 GAGCCCGGGGATGATTTCATCTTGTGGGATGATGTCACAGGCAGGTGCAGACACAGGG 519
    : : : | | | : : : | | | | |
Db 288 GluProGly-----ProAlaGlyValGln-----Gly 296
    : : : | | | : : : | | | | |
QY 518 CCGGCTGTTCTCCACATGAAGAGGTGGACTGTCTCCGGGGACCAAGTTCCATGTAGG 459
    : : : | | | : : : | | | | |
Db 297 ProProGlyProAlaGlyGluGly----- 305
QY 458 CATCAAGACGCTCATGTGCGACAGAGAGCGGCCGTATCAGATCCCGACAGCCATGT 399
    : : : | | | : : : | | | | |
Db 306 -----LysArgGlyAlaArgGlyGluProGlyProSerGlyLeuProGlyProPro--- 322
    : : : | | | : : : | | | | |
QY 398 GCAGCATCCAGCGGTGCTCTGCTTGAAGTTCATCCACACGTGCCAGCGGGCTCCATCA 339
    : : : | | | : : : | | | | |
Db 323 -----GlyGluArgGlyGlyProGly 329
QY 338 GGTAGAAACATCTGGGTGCTGCCCAAAAGCTGCCCAACAAAGAGAGCCAGAGCGCC 279
    : : : | | | : : : | | | | |
Db 330 SerArgGlyPheProGlyAlaAspGlyValAlaGlyPro-----LysGlyProAlaGly 347
    : : : | | | : : : | | | | |
QY 278 AGGAGACAGAACCCAGCAGTGCATGCGCTCGGCTGTGCCTTCATAGACAGGAGCTGA 219
    : : : | | | : : : | | | | |
Db 348 GluArgGlySerProGly----- 353
QY 218 TGTGTGGCTGTACATGTGGAAGATAGAGCCAGATGCCATCTGGGAACCCAGAAACA 159
    : : : | | | : : : | | | | |
Db 354 -----ProAlaGlyProLysGlySerProGlyGlu 363
QY 158 GCAGAGGTTTCATTTTTTTAGCAGTACATGCTGAAGTGAAGACCTTGTGGCTTG 99
    : : : | | | : : : | | | | |
Db 364 Ala-----GlyArgProGlyGluAla 370
QY 98 GGCCTTCCTGCTGCTTTTGAAGGCTCACTGTAGCGGGCCATGCGCTTTTCTCCTCA 39
    : : : | | | : : : | | | | |
Db 371 GlyLeuProGlyAla---LysGlyLeuThrGlySerProGly-----SerProGly 386
    : : : | | | : : : | | | | |
QY 38 CCC 36
Db 387 Pro 387

```

Search completed: May 6, 2004, 11:07:20
Job time : 63.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:52:38 ; Search time 23.5 seconds
(without alignments)

9053.560 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattcattgtgtgggta.....tgggattcctaaaaaaa 2043

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Fgapop 6.0			7.0
Delop 6.0			7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2/1/USPTO_spool/p/US09645078/runat_06052004_104559_21727/app_query.fasta_1.2183
-DB=SwissProt 42 -QFMT=fast -SUFFIX=rs -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=PTO -NORM=ext -HSPSIZE=500 -MINLEN=0 -WAXLEN=200000000
-USER=US09645078 -CGN 1_22 @runat_06052004_104559_21727 -NCFU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500.5	13.4	458	1	C6ST CHICK
2	147	4.0	660	1	YHL1_EBV
3	147	4.0	1453	1	CA11_MOUSE
4	145	3.9	1464	1	CA11_HUMAN
5	143.5	3.8	1172	1	TSP2_HUMAN
6	141	3.8	1460	1	CA11_CANFA
7	140.5	3.8	1003	1	MBD6_HUMAN
8	139.5	3.8	1464	1	CA13_MOUSE
9	138.5	3.7	1372	1	CA21_MOUSE
10	138.5	3.7	1418	1	CA12_HUMAN
11	138	3.7	671	1	CA11_RAT
12	138	3.7	1453	1	CA11_CHICK
13	137.5	3.7	1233	1	MUSA_HUMAN
14	137.5	3.7	1496	1	CA25_HUMAN
15	137	3.7	1366	1	CA21_CANFA
16	136	3.6	830	1	SREC_HUMAN
17	136	3.7	1049	1	CA13_BOVIN
18	135	3.6	784	1	YAV2_XANCV

C 19	134	3.6	1964	1	NTC4_MOUSE
C 20	133.5	3.6	1459	1	CA12_MOUSE
C 21	133	3.6	825	1	ICP0_HSV2H
C 22	133	3.6	1372	1	CA21_RAT
C 23	132.5	3.6	1355	1	CA21_CANCA
C 24	132.5	3.6	2944	1	CA17_HUMAN
C 25	131.5	3.6	1364	1	CA21_BOVIN
C 26	131	3.5	636	1	CA13_RAT
C 27	131	3.5	2003	1	NTC4_HUMAN
C 28	130.5	3.5	1700	1	BAR3_CHITE
C 29	130.5	3.5	2003	1	NTC4_HUMAN
C 30	130	3.5	1175	1	HGN4_RABIT
C 31	128	3.5	1362	1	CA21_CHICK
C 32	127	3.4	830	1	SREC_HUMAN
C 33	127	3.4	1516	1	CA1H_HUMAN
C 34	126	3.4	1712	1	CA24_HUMAN
C 35	125.5	3.4	210	1	K247_HUMAN
C 36	125.5	3.4	316	1	CC07_CAEEL
C 37	125.5	3.4	964	1	IP2_COREF
C 38	125	3.4	680	1	CA1A_HUMAN
C 39	125	3.4	1366	1	CA21_HUMAN
C 40	125	3.4	1466	1	CA13_HUMAN
C 41	124	3.4	1033	1	IP2_STRCO
C 42	124	3.4	2318	1	NTC3_MOUSE
C 43	123	3.3	1046	1	IP2_STRAW
C 44	122.5	3.3	1804	1	CA1B_MOUSE
C 45	122	3.3	627	1	SPD2_NEPLC

ALIGNMENTS

RESULT 1
C6ST_CHICK STANDARD; PRT; 458 AA.

AC Q92179; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).

OS Gallus gallus (Chicken)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

OX NCBI_TaxID=9031;

RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.

RC STRAIN=White leghorn; TISSUE=Embryonic chondrocytes;

RX MEDLINE=95355490; PubMed=7629189; Kato M., Kimata K.,

RA Fukuta M., Uchimura K., Nakashima K.,

RA Shinomura T., Habuchi O.;

RT "Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase."

RL J. Biol. Chem. 270:18575-18580(1995).

CC -1- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN

CC -1- SULFATE.

CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin = adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D49915; BAA08655.1; -

DR PIR; A57397; A57397.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransfer; 1.

KW Transferase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.

FT DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 24 37 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 38 458 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 62 62 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 73 73 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 236 236 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 458 AA; 52253 MW; C9A3B7D0A5086F0C CRC64;
 Alignment Scores:
 Pred. No.: 4,39e-29 Length: 458
 Score: 500.50 Matches: 120
 Percent Similarity: 54.02% Conservative: 75
 Best Local Similarity: 33.24% Mismatches: 133
 Query Match: 13.38% Indels: 33
 DB: 1 Gaps: 13
 US-09-645-078-1 (1-2043) x C6ST_CHICK (1-458)
 QY 227 CTGTCTATGAGGACGAGCCGACGATGCTGTGCTTCTGCTTCTCTCTGCGCTCT 286
 DB 104 LeuGlyIleAlaProGluProArgHisValLeuMetAlaThrArgThr 123
 QY 287 GGCCTCTCTTTGGGGGACGCTTTTGGGAGCACCACCATGTTTCTACCTGATGGAG 346
 DB 124 GlySerSerPheValGlyGluPhePheAsnGlnGlyAsnIlePheTyrLeuPheGlu 143
 QY 347 CCGCCTGGCAGCTG-----TGGATGACCTTCAAGCAGAGCAGCCGCTGGATGCTGCAC 400
 DB 144 ProLeuTrpHisIleGluArgThrValThrPheGluProGlyGlyAlaAsnAlaValGly 163
 QY 401 ATGCTGTG-----CGGATCTGATACGGCGCGCTCTTCTGCGCAGCATGAGCGCTTT 454
 DB 164 SerAlaLeuValTyrArgAspValLeuGlnGlnLeuLeuLeuLeuLeuLeuLeu 183
 QY 455 GATCCCTACATGGAACCTGTGCTCCCGGAGCAG-----TCCAGCCTCTTCAGTGGAG 508
 DB 184 GluSerPheIleSerProAlaProGluGluHisLeuThrAlaAlaLeuPheArgArgGly 203
 QY 509 AACAGCGCGGCGCTGTCTGTCACCTGCTGCTGTCATCATCCACAA---GATGAATC 565
 DB 204 SerSerHisSerLeuLeuGluProValCys-----ThrProSerLeuLeuLysVal 221
 QY 566 ATCCCCGGGCTCACTCAGG---CTCTGTGCACTCAACAGCCCTTGTGAGTGGAG 622
 DB 222 PheGluLysTyrHisCysLysAsnArgArgCysGlyProLeuAsnIleThrLeuAlaAla 241
 QY 623 AAGCGCTGCCCTCTACAGCAGCTGTGCTCAAGGAGGTGCGCTTCTTCAACCTGCAG 682
 DB 242 GluAlaCysArgArgLysGlnHisMetAlaLeuLysThrValArgIleArgGlnLeuGlu 261
 QY 693 TCCCTCTACCGCTGTCTGAAGAGCCCTCCCTCAACCTGCATATCGTGCACCTGTCGG 742
 DB 262 PheLeuGlnProLeuAlaGluAspProArgLeuAspLeuArgIleIleGlnLeuValArg 281
 QY 743 GACCCCGGGCGCTGCTCCGTTCCTCCGAGACGACCAAGGAGATCTCATGATTGACAGT 802
 DB 282 AspProArgAlaValLeuValSerArg-----MetVal----- 292
 QY 803 CGCATTTGATGGGGCAGCATGACCAAACTCAAG----- 838
 DB 293 ---AlaPheSerGlyLysTyrGluSerTrpAlaAlaGluGlyGluAlaPro 311
 QY 839 ---AAGGAGGACCAACCTACTATGTGATGAGGTGATCGTCCCAAGG---CAGCTGGAG 892
 DB 312 LeuGlnGluAspGlu---ValGlnArgLeuArgGlyAsnCysGluSerIleArgLeuSer 330
 QY 893 ATCTACAAAGCATCCAGTCTCTGCCCAGGCCCTGACGAGAACCTACCTGCTGTGCGC 952
 DB 331 AlaGluLeuGlyLeuArgGln---ProArgTrpLeuArgGlyArgTyrMetLeuValArg 349

QY 953 TATGAGGACCTGGCTCGAGCCCTGTGCGCCAGACTTCCCGAATGTATGTAATTCGTGGGA 1012
 DB 350 TyrGluAspValAlaArgAlaProLeuArgLysAlaLeuGluMetTyrArgPheAlaGly 369
 QY 1013 TTGGAATCTTGGCCCATCTCTCAGACCTGGTGCATTAACATCACCAGGCAAGGCATG 1072
 DB 370 IleHisProThrProGlnValGluIleArgAlaAsnThrGlnAlaPro---Gln 388
 QY 1073 GGTGACACACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCTCCAGGCTTGGCGC 1132
 DB 389 AspSerAsnGlyIleTyrSerThrGlnLysAsnSerSerGluGlnPheGluLysTrpArg 408
 QY 1133 TGGTCTTTGGCCTATGAAAGCTTCTCGACTTCAGAAAGCCTGTCGCGATGCGATGAAT 1192
 DB 409 PheSerIleProPheLysLeuAlaGlnValValGlnAspAlaCysGluProAlaMetArg 428
 QY 1193 TTGCTGGCTACCGCCACGCTCAGATCTGAACAAGAACAGAGAAACCTGTTGCTGGATCTT 1252
 DB 429 LeuPheGlyTyrLysLeuAlaSerSerAlaGlnGluLeuThrAsnArgSerLeuSerLeu 448
 QY 1253 CTG 1255
 DB 449 Leu 449

RESULT 2

YHLL_EBV
 ID YHLL_EBV STANDARD; PRT; 660 AA.
 AC P03181;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical BHLF1 protein.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tufnell P.S., Ward R., White P., Young I.P., Zannis-Haber V.,
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."
 RL Nature 310:207-211 (1984).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: V01555; -; NOT_ANNOTATED_CDS.
 DR PIR: A03742; Q0833.
 KW Hypothetical protein; Early protein; Repeat.
 FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 SQ SEQUENCE 660 AA; 56244 MW; 86D1D67A37152A2 CRC64;

Alignment Scores:

Pred. No.: 0.0053 Length: 660
 Score: 147.00 Matches: 122
 Percent Similarity: 28.78% Conservative: 19
 Best Local Similarity: 24.90% Mismatches: 175
 Query Match: 3.97% Indels: 174
 DB: 1 Gaps: 29

US-09-645-078-1 (1-2043) x YHLL_EBV (1-660)

RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RL of the 3'-untranslated region."
CC Nucleic Acids Res. 16:773-773(1988).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- SIMILARITY: Contains 1 VWF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: U08020; AAA88912.1; -
CC EMBL: X15896; CAA33904.1; -
CC EMBL: M14823; AAA37333.1; -
CC EMBL: M17491; AAA37334.1; -
CC EMBL: X06753; CAA29927.1; -
CC EMBL: K03036; AAA37332.1; -
CC EMBL: K03029; AAA37332.1; JOINED.
CC EMBL: K03030; AAA37332.1; JOINED.
CC EMBL: K03031; AAA37332.1; JOINED.
CC EMBL: K03032; AAA37332.1; JOINED.
CC EMBL: K03033; AAA37332.1; JOINED.
CC EMBL: K03034; AAA37332.1; JOINED.
CC EMBL: K03035; AAA37332.1; JOINED.
CC PIR: S57243; S21626.
CC MGI: MGI:88467; Colla1.
CC InterPro: IPR008161; C1q helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib collagen_C.
DR InterPro: IPR002181; Fibrinogen_C.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COL1; 1.
DR Pfam: PF01391; Collagen; 18.
DR ProDom: PD000007; C1q_helix; 1.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COL1; 1.
DR SMART: SM00214; VWF; 1.
DR PROSITE: PS01208; VWF_C_1; 1.
DR PROSITE: PS0184; VWF_C_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 VWF.
FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 168 1181 TRIPLE-HELICAL REGION (C-TERMINAL).
FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 1450 1450 A -> V (IN REF. 5).
SQ SEQUENCE 1453 AA; 137944 MW; 3B802E535DF81808 CRC64;

Alignment Scores:
Pred. No.: 0.0672 Length: 1453
Score: 147.00 Matches: 104
Percent Similarity: 29.98% Conservative: 21
Best Local Similarity: 24.94% Mismatches: 136
Query Match: 3.97% Indels: 157

DB: 1 Gaps: 24
US-09-645-078-1 (1-2043) x CALL_MOUSE (1-1453)
QY 1144 GGGCAAGACACAGCCGACGCTGGAGACATTAAAGGCATCCCTGGCAT-----TTG 1092
DB 231 GlyGluArgGlyProGlyProGlnGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 250
QY 1091 TGTGAAAGCGTGTGTAC----- 1074
DB 251 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGly 270
QY 1073 CATGCCCTTCCTCGGCTGATGTATGCACCCAGGTCTCAAGATGGGCAAGATTCCA 1014
DB 271 ProAlaGlyProLysGlyGluProGlySerPro---GlyGluAsnGlyAlaPro----- 287
QY 1013 ATCCCAAGATTATACATTCGGAAGTCTGGGACACAGGGGCTC----- 969
DB 288 -----GlyGlnMetGlyProArgGlyLeuProGlyGluArgGly 300
QY 968 GAGCCAGCTCT-----CATAGCGCACAGCAGGTAGC 936
DB 301 ArgProGlyProProGlyThrAlaGlyAlaArgGlyAsnAspGlyAlaValGlyAlaAla 320
QY 935 GTTCCTGCAGGCGCTTGGGCAAGACCTGGATGCTTGTAGATCTCCAGCTGGCTTTGGC 876
DB 321 GlyProProGlyProThrGlyProThrGly-----ProProGlyPhePro 335
QY 875 AGATGACCTGCATCATAGTAGGTGGTCTCT-----CCT 840
DB 336 GlyAlaValGlyAlaLysGlyGluAlaGlyProGlnGlyAlaArgLysGlyGluGlyPro 355
QY 839 TCTTGAGTTTTTGTCTCATGCTGCCCCCATCACATCGCACTGTCAATCATGAGATCTCCCT 780
DB 356 GlnGlyValArgGlyGluProGlyProGly-----Pro 367
QY 779 TTGTCGTTCTCGGAACGGAACAGCGCCGCGGGTCCCGA-----CAGGTGACAGTAT 723
DB 368 AlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAlaLysGly 387
QY 722 GCAGGTTGAGGAGGGGCTTTACGACG-----GGTAGAGGACTGCA 678
DB 388 AlaAsnGlyAlaProGlyLeuAlaGlyAlaProGlyPheProGlyAlaArgGlyProSer 407
QY 677 GGTGAAGA-----AGCGCACCTCTTTAGCACACACCTGCTGTAGAGCGGCGAGGCT 624
DB 408 GlyProGlnGlyProSerGlyProPro---GlyProLysGly---AsnSerGlyGluPro 425
QY 623 TCTCCACACCTCAAGAGGCTGTGCTGACACAGGAGCTGCAGTACGCCGCGGGGATGA 564
DB 426 GlyAlaProGlyAsnLysGly-----AspThrGlyAlaLysGlyGluProGlyAla--- 442
QY 563 TTTTCATCTTGTGGATGATGTACAGGCGAGGTGCAGAACACAGGCGCGGCTGTCTCCC 504
DB 443 -----ThrGly----- 444
QY 503 ACTGAAAGGCTGACTGTCTCCGGGACACAGGTTCTCATGTAGGCATCAA---AGAGCG 447
DB 445 -----ValGlnGlyProProGlyProAlaGlyGluGlnGlyLysArg 458
QY 446 TCATGTCGCACAAGAAGACGGCCCTATCAGATCCCGCACAGCCATGTGCAGCATCCAGG 387
DB 459 GlyAlaArgGlyGluProGlyProSerGlyLeuProGlyProPro----- 473
QY 386 CGGTGCTCTGCTGAAGGTTCATCCACAGTGCAGCGCGGCTCCATCAGGTAGAAAACAT 327
DB 474 -----GlyGluArgGlyGlyProGlySerArgGlyPhe 484
QY 326 CTGGGTGCTGCCAAAAAGCTCCCCACAAAAGAGAGCCAGAGCGCCAGGAGACAGAA 267
DB 485 ProGlyAlaAspGlyValAlaGlyPro-----LysGlyProSerGlyGluArgGlyAla 502
QY 266 CCAGCAGTGCATCGCGCTCGGCTGTGCTTCATAGACAGGAGCTGATGTTGTGGCTGT 207

Db 503 Progly-----504
 QY 206 ACATGTGGAAGATAGACCAAGATGCCATCTGGGAACACAGAAACAGAGAGCTTCA 147
 Db 505 -----ProAlaglyProLysGlySerProGlyGluAla-----515
 QY 146 TTTTITTAGGCAGTAGCATTCGTCAGTAGTGAAGACCTTGTGGCTTGGCTTCCCTGCT 87
 Db 516 -----GlyArgProGlyGluAlaglyLeuProGly 525
 QY 86 GCTTTTGAAGAGCTCACTGCTAGCGGCCATCGCTTTTCTCTCTACCC 36
 Db 526 Ala---LysGlyLeuThrGlySerProgly-----SerProglyPro 538
 RESULT 4
 CALL HUMAN
 ID CALL HUMAN STANDARD; PRT; 1464 AA.
 AC P02452; Q14037; Q15176;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-472 FROM N.A.
 RX MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
 RA Jaenisch R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
 RT chain of human type I procollagen.";
 RL Biochem. J. 253:919-922(1988).
 RN [2]
 RP SEQUENCE OF 1-181 FROM N.A.
 RX MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 RA Myers J., Williams C., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons.";
 RL Nature 310:337-340(1984).
 RN [3]
 RP SEQUENCE OF 162-301.
 RX TISSUE=Skin;
 RC MEDLINE=71038625; PubMed=5529814;
 RA Click E.M., Bornstein P.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706(1970).
 RN [4]
 RP SEQUENCE OF 263-268.
 RX TISSUE=Skin;
 RC MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected
 RT vertebrate collagens. A possible role of the carbohydrate in fibril
 RT formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [5]
 RP SEQUENCE OF 425-1464 FROM N.A.
 RX MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [6]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 RX TISSUE=Bone;
 RC

RX MEDLINE=88124208; PubMed=3340531;
 RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [7]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to
 RT transcriptional control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [8]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 RT Promoter structure, AluI repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [9]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.J.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans.";
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RL cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RN Hum. Mutat. 9:300-315(1997).
 RP REVIEW ON OI VARIANTS.
 RX MEDLINE=91374476; PubMed=1895312;
 RA Byers P.H., Wallis G.A., Willing M.C.;
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
 RL J. Med. Genet. 28:433-442(1991).
 RN [13]
 RP REVIEW ON OI VARIANTS.
 RX MEDLINE=97169389; PubMed=9016532;
 RA Dalgleish R.;
 RT "The human type I collagen mutation database.";
 RL Nucleic Acids Res. 25:181-187(1997).
 RN [14]
 RP VARIANT OI-II CYS-1166.
 RX MEDLINE=86287390; PubMed=3016737;
 RA Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
 RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
 RT change in one human pro alpha 1(I) collagen allele.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
 RN [15]
 RP VARIANT OI-II ARG-569.
 RX MEDLINE=87222295; PubMed=3108247;
 RA Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
 RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
 RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
 RL collagen.";
 RN J. Biol. Chem. 262:7021-7027(1987).
 RN [16]
 RP VARIANT OI-II CYS-926.
 RX MEDLINE=88033031; PubMed=3667599;
 RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;

Db 683 yYrAlaGlyAspGlyLeuIleCysGlyGluAspSerAspLeuAspGlyTrpProAsnLe 703
 Qy 1564 CAGAGACTTTGGCGCTGGAGGCTATTAAACACAGACACAGATATCACTGGAATGATCCA 1623
 Db 703 uAsnLeuValCysAlaThrAsnAlaThrTyHis-----CysII 716
 Qy 1624 TAAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCTTTCCACAAAGACTCAC 1683
 Db 716 eLyAspAsnCysProHisLeuProAsnSerGlyGlnGluAspPheAspLysAspGlyII 736
 Qy 1684 AGCATTTTCCACAGAG----- 1699
 Db 736 eGlyAspAlaCysAspAspAspAspAspAspGlyValThrAspGluLysAspAsnCy 756
 Qy 1700 -ATGCAATCTGAGCCCTTGAGTCCCAATGGATTCAAGGAAGAGTGGGAACAAG 1758
 Db 756 sGlnLeuLeuPheAsnProAsgGlnAlaAspTyAsp---LysAspGluValGlyAspAr 775
 Qy 1759 GTTGATGCCCTACTTATGAGTTCGACCATAGCTATCGGTAATCAGAAATA 1810
 Db 775 gCysAspAsnCys-----ProTyrValHisAsnProAlaGlnIle 788
 RESULT 6
 CALL CANFA STANDARD; PRT; 1460 AA.
 ID CALL CANFA STANDARD; PRT; 1460 AA.
 AC Q9XSJ7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
 RC TISSUE=Skin;
 RX MEDLINE=21023337; PubMed=11147834;
 RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
 RT "Sequence of normal canine COL1A1 cDNA and identification of a
 RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case
 RT of canine osteogenesis imperfecta";
 RL Arch. Biochem. Biophys. 384:37-46(2000).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 CC (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
 CC (OI).
 CC -!- SIMILARITY: Contains 1 WFCC domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF153062; AAD34619.1; -
 CC InterPro; IPR008161; C1g helix.
 CC InterPro; IPR008160; Collagen.
 CC InterPro; IPR000885; Fib collagen_C.
 CC InterPro; IPR002181; Fibrinogen_C.
 CC InterPro; IPR001007; WFC_C.
 CC Pfam; PF01410; COLFI; 1.
 CC Pfam; PF01391; Collagen; 18.
 CC ProDom; PD000007; C1g helix; 2.
 CC ProDom; PD002078; Fib collagen_C; 1.
 CC SMART; SM00038; COLFI; 1.
 CC SMART; SM00214; WVC; 1.

DR PROSITE; PS01208; WFCC_1; 1.
 DR PROSITE; PS0184; WFCC_2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Disease mutation.
 FT SIGNAL 1 22
 FT PROPEP 23 157
 FT CHAIN 158 1214
 FT PROPEP 1215 1460
 FT CHAIN 158 1214
 FT PROPEP 1215 1460
 FT DOMAIN 34 92
 FT DOMAIN 158 174
 FT DOMAIN 175 1198
 FT DOMAIN 1199 1214
 FT SITE 741 743
 FT SITE 1089 1091
 FT CARBOHYD 1361 1361
 FT VARIANT 208 208
 FT SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;
 SQ
 Alignment Scores:
 Pred. No.: 0.0186 Length: 1460
 Score: 141.00 Matches: 102
 Percent Similarity: 30.68% Conservative: 25
 Best Local Similarity: 24.64% Mismatches: 137
 Query Match: 3.81% Indels: 151
 DB: 1 Gaps: 24
 US-09-645-078-1 (1-2043) x CALL CANFA (1-1460)
 Qy 1144 GGGCAAGACAGCGCCAGACCTGGAGACATTAAGGGCATCCCTGGCAT-----TTG 1092
 Db 238 GlyGluArgGlyProGlyProGlyGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 257
 Qy 1091 TGTGGAAACGCTGGTCAC----- 1074
 Db 258 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAspAlaGly 277
 Qy 1073 CCATGCCCTTCCTCGGCTGATGTTATGCACCCAGCTCTGAAGATGGGCGCAAGATTCCA 1014
 Db 278 ProAlaGlyProLysGlyGluProGlySerPro---GlyGluAsnGlyAlaPro----- 294
 Qy 1013 ATCCCACGAATTCATATCGGGAAGTCTGGCCACAGGGCTC----- 969
 Db 295 -----GlyGlnMetGlyProArgGlyLeuProGlyGluArgGly 307
 Qy 968 -----GAGCCAGTCTCTCATACGCGCACAA-----GCAGGTAGCTTCTCTGCA 927
 Db 308 ArgProGlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAlaAla 327
 Qy 926 GGGCCTTGGCAAGAGTGGATGGTCTTGTAGATCTCCAGCTGGCTTGGCAGATGACCT 867
 Db 328 GlyProGlyProThrGlyProAla-----GlyProGlyPheProGlyAlaVal 345
 Qy 866 GCATCATAGTAGGCTTGGTCT-----CCTTCTTGAGTT 831
 Db 346 GlyAlaLysGlyGluAlaGlyProGlnGlyAlaArgGlySerGluGlyProGlnGlyVal 365
 Qy 830 TTTGCTCATGCTCCCTCCATCACAATCGACTGTCAATCATGAGATCTCCCTTTGCGGT 771
 Db 366 ArgGlyGluProGlyProGly-----ProAlaGlyAla 377
 Qy 770 CTGGGACGGAACACAGGCGCGGGTCCCGG-----CCAGGTGCAGATATGCAAGTTGA 714
 Db 378 AlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAlaLysGlyAlaAsnGly 397
 Qy 713 GGGAGGGGCTTTTCACAGCG-----GGTAGAGGGACTCGAGTTGAAGA 669
 Db 398 AlaProGlyIleAlaGlyAlaProGlyPheProGlyAlaArgGlyProSerGlyProGln 417
 Qy 668 -----AGCCACCTCTTCCAGCACACACGTGGCTGTAGAGCGCGAGGCTTCTCCACCA 615
 Db 418 GlyProSerGlyProPro---GlyProLysGly---AsnSerGlyGluProGlyAlaPro 435
 Qy 614 CCTCAAGGGCTGTTGACTGCACAGAGCCTGCAGTGAGCCCGGGGATGATTTTCATCTT 555

```

Db 436 GlyAsnLysGly-----AspThr-GlyAlaLysGlyGluProGly----- 448
QY 554 GTGGGATGATGTACAGGAGGTCAGAAACACAGAGGCGCGGCTGTTCTCCCACTGAAGA 495
Db 449 -----ProThr----- 450
QY 494 GCCTGGAGTCTCTCGGGACACAGGTTCCATGTAGGATCAA-----AGACGTCATGTCG 438
Db 451 -----GlyIleGlnGlyProProGlyProAlaGlyGluGlyLysArgGlyAlaArg 468
QY 437 ACAAGAACAGCGCGCGTATCAGATCCCGCACAGCCATGTGCAGCATCCAGCGGTGCTCT 378
Db 469 GlyGluProGlyProThrGlyLeuProGlyPro----- 480
QY 377 GCTTGAAGTCAATCCACAGTCGACAGCGCGGCTCCATCAGTAGAACAATCTGGTGTCT 318
Db 481 -----GlyGluArgGlyGlyProGlySerArgGlyPheProGlyAla 494
QY 317 GCCCAAAAGCTGCCCAACAAAGAGAGAGCCAGAGCGCCAGAGACAGAACACGACGCT 258
Db 495 AspGlyValAlaGlyPro-----LysGlyProAlaGlyGluArgGlySerProGly--- 511
QY 257 GCATCGCTCGGGCTGTGCTTCATAGACAGGAGCTGATGTTGTGCTGTACATGTGA 198
Db 511 ----- 511
QY 197 AGAATAGAGCAAGATGCGCTCTCGGAAACCCAGAAACAGAGGCTTCATTTTGTAG 138
Db 512 -----ProAlaGlyProLysGlySerProGlyGluAla----- 522
QY 137 GCAGTAGCATGTGTGTAAGTGAAGACCTGTGTGGCTTGGCTCCCTGCTGCTTTTGA 78
Db 523 -----GlyArgProGlyGluAlaGlyLeuProGlyAla-----Lys 534
QY 77 AGGCTCACTGCTAGCGCGGCGCATGCGCTTCTCTCTCTACCC 36
Db 535 GlyLeuThrGlySerProGly-----SerProGlyPro 545

RESULT 7
MBD6 HUMAN
ID MBD6 HUMAN STANDARD; PRT; 1003 AA.
AC Q96N6; Q8N3M0; Q8N81; Q96Q00;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Methyl-CpG binding domain protein 6.
GN MBD6 OR KIAA1887.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagaetsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
```

```

RL DNA Res. 8:179-187(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=AmnYGdala;
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK056399; BAB71176.1; -
CC EMBL; AK093078; BAC04045.1; -
CC EMBL; AB067474; BAB67780.1; -
CC EMBL; AL834230; CAD38908.1; -
CC Genew; HGNC:20445; MBD6.
CC InterPro; IPR001739; Methyl-CpG_bind.
CC Pfam; PF01429; MBD; 1.
CC SMART; SM00391; MBD; 1.
CC DOMAIN 17 61
CC FT DOMAIN 140 846
CC FT CONFLICT 637 637 D -> G (IN REF. 1; BAB71176).
CC FT CONFLICT 802 802 Q -> QQ (IN REF. 1; BAC04045).
CC FT CONFLICT 802 802 Q -> QQ (IN REF. 1; BAC04045).
SQ SEQUENCE 1003 AA; 101200 MW; 6C8B8693AA6A3BE6 CRC64;

Alignment Scores:
Pred. No.: 0.081 Length: 1003
Score: 140.50 Matches: 137
Percent Similarity: 32.11% Conservative: 38
Best Local Similarity: 25.14% Mismatches: 156
Query Match: 3.75% Indels: 216
DB: 1 Gaps: 30

US-09-645-078-1 (1-2043) x MBD6_HUMAN (1-1003)
QY 64 CTAGCAGTGAAGCTCTCAAAAGCAGGAGGAG-----CCNAGCCACAGGT 111
Db 319 LeuAlaSerSerLeuLeuSerAlaAlaLysAlaGlnHisProLeuProProPro 338
QY 112 CTTCCACTTCAGCACAACTACTGCTCTAAATAAATGAAGCTCTCTTCTGTTTC 171
Db 339 SerThrLeuGlnGlyArgArgPro-----ArgAlaGlnAlaProSerAlaSer 354
QY 172 CCAGATGGCCATCTTGGCTCTATTCTTCACATGTACAGCCACACATGCTCCCT--- 228
Db 355 -----HisSerSerLeuArgProSerGlnArgArgProArgProThr 371
QY 229 -----GTCTATGAGGCACAGCC-----CGAGCCGATGCACTGCT 264
Db 372 ValPheArgLeuLeuGlyArgGlyProGlnThrProArgArgSerArgProArgAla 391
QY 265 GGTCTCTGT-----CTTCTGCG 281
Db 392 ProAlaProValProGlnProPheSerLeuProGluProSerGlnProLeuProSer 411
QY 282 GCTCTGCTCTCTCTT-----TTGTGGGGCGAGCTTT--- 311
Db 412 ValLeuSerLeuGlyLeuProThrProGlyProSerHisSerAspGlySerPheAsn 431
QY 312 ---TTGGGCGAGCCAGATGTTTCTACCTGATGAGCCCGCTCGACGCTGGATGA 368
Db 432 LeuLeuGlySerAspAlaHisLeu-----ProProProThrLeu----- 445
QY 369 CTTTCAAGCAGACACCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 428
Db 446 ---SerSerGlySerProPro----- 451
QY 429 TCTTCTTGTGCGACATGAGCGTCTTTGATGCTTACATGGAACCTGCTCCCGGAGACGT 488
```

Db 452 -----GlnProArgHisProIleGlnProSerProGlyThr 465
QY 489 CCAGCCTTTTCACT-----GGGAGACACCGGGCCCTGTCTGCACCTGCCT 539
Db 466 SerGlySerLeuSerSerValProGlyAlaProAlaProAlaSerLysAlaPro 485
QY 540 GTGACATCATCCCAAGATGAATCATCCCCC-----GGGCTCACTGCAGGCTCC 590
Db 486 ValValProSerProValLeuGlnSerProSerGluGlyLeuGlyMetGlyAlaGlyPro 505
QY 591 TGTGAGCTCAACAGCCCTTGTAGGTGGTGGAGAGGCTGCCCTCTACAGCCACCTGG 650
Db 506 AlaCysProLeuProProLeu-AlaGlyGlyGluAlaPheProPheProSerProGlu-- 524
QY 651 TGCTCAAGAGGTGGCTT-----CTTCAAGCTGCAGTCCCTCT 689
Db 525 ----GlnGlyLeuAlaLeuSerGlyAlaGlyPheProGlyMetLeuGlyAlaLeuProLe 543
QY 690 ACCC----- 693
Db 543 uProLeuSerLeuGlyGlnProProProSerProLeuLeuAsnHisSerLeuPheGlyVa 563
QY 694 -----GCTGCTGAAGACCCCTCCCTCAACCTGCATATCGTCACCT---GGTCC 740
Db 563 lLeuThrGlyGlyGlnProProProGlnProLeuLeuProProProGlyGlyPr 583
QY 741 GGGACCCCGGGCGGTGTCCGTCCCGAGAACGACAAAGGAGATCTCATGATTGACA 800
Db 583 oGlyProPro----- 586
QY 801 GTGCGATTGTGATGGGCGAGCATGAGCAAAACTCAAGAGGAGCAACCCCTACTATG 860
Db 587 -----LeuAlaProGlyGluProGluGlyProSerLeuLeuVa 599
QY 861 TGATCAGGTGATCTGCCAAGCCAGCTGGAGATCTCAAGACCAT----- 906
Db 599 l---AlaSerLeuLeuProProProSerAspLeuLeuProProProSerAlaProPr 618
QY 907 -----CCAGTCTCTGCCAAGCCCT---CGAGAACGCT 938
Db 618 oSerAsnLeuLeuAlaSerPheLeuProLeuLeuAlaLeuGlyProThrAlaGlyAsp-- 637
QY 939 ACCTGCTTTGTGGCTATGAGGACCTGGCTCGAGC-----CCCTGTGCCCC 983
Db 638 -----GlyGluGlySerAlaGluGlyAlaGlyGlyProSerGlyG1 651
QY 984 AGACTTCCGATGTATGATTCGTGGATGGATTTCTTCCCA----- 1029
Db 651 uProPhe-----SerGlyLeuGlyAspLeuSerProLeuLeuPheProPr 666
QY 1030 -----TCTTCAGACCTGGGTGCATTAACATCA 1055
Db 666 oLeuSerAlaProProThrLeuLeuAlaLeuAsnSerAlaLeuLeuAlaThrLeuAs 686
QY 1086 CCGAGGACAGGCGATGGTGACACGCTTCCACAAATGCCAGGATGCCCTTATG 1115
Db 686 pProSerGlyThrPro-----ProGlnProCys----- 696
QY 1116 TCTCCAGGCTTGGGCTGGTCTTTCCTATGAAAGGTTTCTCGACTTCA----- 1167
Db 697 -----ValLeuSerAlaProGlnProGlyProProThrSerSerValTh 711
QY 1168 -----GAAAGCTGTGGCGATGCCATGAATTTGCTGGGCT 1202
Db 711 rThrAlaThrAspProGlyAlaSerSerLeuGlyLysAlaProSerAsnSerGlyAr 731
QY 1203 ACCGCCAGCTCAGATCTCAAGAACAGACAGAACTCTGTGATCTTCTGCTACCT 1262
Db 731 gProGlnLeuLeu-----SerProLeuLeuGlyAlaSerLeuLeuG1 746
QY 1263 GGAAGTGTGAGCAATCCATCAAGAGGTTGAGAGGCTTGTGTCACCACTGG 1318

Db 746 yAspLeuSerSerLeuThrSerSerProGlyAlaLeuProSerLeuLeuGlnProProGl 766
QY 1319 TGTTCAGCTCTCACT 1331
Db 766 yProLeuLeuSer 770
RESULT 8
CA13 MOUSE
ID CA13 MOUSE STANDARD; PRT; 1464 AA.
AC P08121; Q61429; O9CRN7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
complete DNA sequence.";
RL Gene 147:161-168(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Richards S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Villalón D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahney J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Rodriguez R.W., Touchman J.W., Green E.D., Dickson M.C.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogeli G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
alpha-1 type-III collagen chain.";
RL Gene 61:225-230(1987).
RN [4]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Liao G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
(III) collagen gene.";
RL J. Biol. Chem. 260:3773-3777(1985).
RN [5]
RP SEQUENCE OF 810-1464 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schirali L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni I., Mashima J., Mazzarelli C., Mombartzs P.,
RA Nordone P., Kling B., Ringwald M., Rodriguez I., Sakamoto S.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RP [6]
RP SEQUENCE OF 1442-1464 FROM N.A.
RP STRAIN=CS7BL/6;
RX MEDLINE=91274355; PubMed=2054384;
RA Matsueranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RT collagen mRNAs";
RL Biochim. Biophys. Acta 1089:241-243(1991).
CC -I- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -I- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylysines.
CC -I- PM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains.
CC -I- PM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC the oxygen atom of a post-translationally added hydroxyl group (By
CC similarity).
CC -I- SIMILARITY: Contains 1 VWFC domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X52046; CAA36279.1; --
DR EMBL; BC043089; AAH43089.1; --
DR EMBL; BC058724; AAH58724.1; --
DR EMBL; M18933; AAA37338.1; --
DR EMBL; K03037; -- NOT ANNOTATED_CDS.
DR EMBL; AK019448; BAB31724.1; --
DR EMBL; X57983; CAA41048.1; --
DR PIR; A27353; A27353.
DR PIR; S59856; S59856.
DR MGD; MGI:88453; Clg3a1.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWFC_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SMC0038; COLFI; 1.
DR SMART; SMC0014; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS00184; VWFC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 154
FT CHAIN 155 1203
FT Amino-terminal proppeptide.
FT COLLAGEN ALPHA 1(III) CHAIN.

```

OY 491 TGGACTGCTCCGGGACCCAGGTTCCATGTAGGCATCAAAAGACGCTCATGTCGCACAAGA 432
DB 1003 -----ProGlyThrAlaGlyGluProGly---ArgAspGlyAsnProGlySerAs 1018
OY 431 AGAGCGCCGCTATCAGATCCCGCAGCCATGTCAGCATCCAGCGCGTCTGCTTGA 372
DB 1018 pGlyGlnPro-----GlyArgAspGlySerProGlyGlyGlyAspAr 1033
OY 371 AGGTCATCCACACGCTCCAGCGCGCTCCATCAGGTAGAAACATCTGGGTGCTGCCCAA 312
DB 1033 gGlyGluAsnGlySerProGly--AlaProGlyAlaProGlyHisProGlyProGly 1052
OY 311 AAAGTCGCCCAAAAGACAGGACAGCCAGCCAGGAGACAGACAGACAGCGTGCATGC 252
DB 1053 ProValGlyProSerGlyLysSerGlyAspArgGlyGluThrGlyProAla----- 1069
OY 251 CTCGGGCTGTGCTTCATACACAGGAGGTGTTGCTGCTCATCATGGAAGAATA 192
DB 1069 ----- 1069
OY 191 GAGCAAGATGCGCATCTGGGAAACCCAGAACAGCAGGAGCT 150
DB 1070 -----GlyProSerGlyAlaProGlyProAlaGlyAla 1080

RESULT 9
CA21_MOUSE
ID CA21_MOUSE STANDARD; PRT; 1372 AA.
AC Q01149;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2 OR COLA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Calvaria;
RX MEDLINE=92372043; PubMed=1505972;
RA Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;
RT "Sequence analysis of a full-length cDNA for the murine pro alpha
RT 2(I) collagen chain: comparison of the derived primary structure with
RL human pro alpha 2(I) collagen.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RN [3]
RP SEQUENCE OF 1-110 FROM N.A.

```

```

RC TISSUE=Calvaria;
RX MEDLINE=92084969; PubMed=1748823;
RA Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,
RA Wenstrup R.J.;
RT "Construction of a full-length murine pro alpha 2(I) collagen cDNA by
RT the polymerase chain reaction.";
RL J. Invest. Dermatol. 97:980-984(1991).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RX MEDLINE=87289650; PubMed=3039494;
RA Rossi P., de Crombrughe B.;
RT "Identification of a cell-specific transcriptional enhancer in the
RT first intron of the mouse alpha 2 (type I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X58251; CAA41205.1; -
CC EMBL; BC007158; AAH07158.1; -
CC EMBL; BC042503; AAH42503.2; -
CC EMBL; K01832; AAA37331.1; -
CC F&R; A43291; A43291.
CC MGD; MGI:88468; Colla2.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC ProDom; PD0000007; Clg_helix; 6.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC SMART; SM00038; COLFI; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Collagen; Signal; Pyridone carboxylic acid.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 85 AMINO-TERMINAL PROPEPTIDE
FT CHAIN 86 1108 (BY SIMILARITY)
FT PROPEP 1109 1372 COLLAGEN ALPHA 2(I) CHAIN.
FT MOD_RES 86 86 CARBOXYL-TERMINAL PROPEPTIDE
FT MOD_RES 90 90 PYRROLIDONE CARBOXYLIC ACID (BY
FT MOD_RES 90 90 SIMILARITY).
FT MOD_RES 90 90 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT MOD_RES 90 90 INVOLVED IN CROSS-LINKING
FT MOD_RES 90 90 (BY SIMILARITY).
FT CARBOHYD 1273 1273 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 15 15 V -> A (IN REF. 4).
FT CONFLICT 1167 1167 R -> T (IN REF. 1).
SQ SEQUENCE 1372 AA; 129557 MW; 0D17DF5D6C1452D1 CRC64;

Alignment Scores:
Pred. No.: 0.0279 Length: 1372
Score: 138.50 Matches: 104
Percent Similarity: 30.67% Conservative: 15
Best Local Similarity: 26.80% Mismatches: 134
Query Match: 3.74% Indels: 136
DB: 1 Gaps: 24

US-09-645-078-1 (1-2043) x CA21_MOUSE (1-1372)

```

QY 1127 AAGCTGGAGACATTAAAGGCATCCCTGGCATTTGTGTGGAAAGCGTGTCTACCCATGC 1068
Db LysProGlyGluArg---GlyHisPro---GlyLeuAlaGlyAlaArgGlyAlaProGly 529
QY 1067 CCTGCGCTCGGTGATGTTATGCACCCAGGTCTGAAGATGGGCAAGAAATTCGAATCCA 1008
Db Pro----- 530
QY 1007 CGAATTCATACATTCGGGAAGTCTGGGCCACAGGGCTCGAGCCAGGTCTCTCATAGCGCA 948
Db -----AspGlyAsnAsnGlyAlaGlnGly---ProProGlyProGlnGlyVal 545
QY 947 CAACAGGTAGCTTCTCGAGGGCTTGGGCAAGACTGGATGGTCTTGTAGATCTCCA 888
Db GlnGlyGlyLysGlyGlnGlnGlyPro-----AlaGlyPro 557
QY 887 GCTGCTTTGGCAGATGACCT-----GCATCACATAGTAGGTTGGTCTCTCTCT 837
Db ProGlyPheGlnGlyLeuProGlyProSerGlyThrThrGlyGluValGlyLysProGly 577
QY 836 TGAGTTTGTCTGCTGCTGCCCATCATGCGACTGTCATCATGAGATCTCCCTTG 777
Db -----GluArgGlyLeuPro----- 582
QY 776 TCGCTTCTCGGAACGGAACACGCGCCGGGGTCCCGACAGGCTGCACGATATGCAGGT 717
Db -----GlyGluPheGlyLeuProGlyProAlaGlyPro----- 593
QY 716 TGAGGAGGGTCTTTCAGACGCGGTAGAGGACTGAGGTGAAGA-----AGCGCACT 660
Db -----ArgGly-----GluArgGlyThrProGlyGluSerGlyAlaAlaGly 607
QY 659 CTTGAGCACACAGTGGCTGTAGG-----AGCGGAGGCTTCTCCACCACTCAAGG 606
Db ProSerGlyProIleGlySerArgGlyProSerGlyAlaProGlyProAspGlyAsnLys 627
QY 605 GCTGTTGACTGCACAGGAGCTGCAGTGAGCCCGGGGAGATTCATCTTGTGGGATGA 546
Db GlyGlu-----AlaGlyAlaValGlyAlaProGly----- 637
QY 545 TGTACAGGAGGTCTCAGACAGAGGCGCGCTGTTCTCCACTGAAAGAGGTGGACT 486
Db -----SerAlaGlyAlaSerGlyProGlyGlyLeuProGlyGluArgGlyAlaAl 654
QY 485 G-----TCTCCGGGACACAGTTCATAGTCATCAAA-----GACGCTCATGT 441
Db acGlyileProGlyGlyLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 674
QY 440 CGCAAGAGACGCGCGGTATCAGATCCGCACAGCCATGTCAGCATCCAGCGGGTGC 381
Db nThrGlyArgAspGlyAlaArgGlyileProGlyAla---ValGlyAlaProGlyProAl 693
QY 380 TCTGCTTGAAGTTCAT-----CCACAGTGCAGCGGGCT 345
Db aglyAlaSerGlyAspArgGlyGluAlaGlyAlaAlaGlyProSerGlyProAlaGlyPr 713
QY 344 -----CATCAGGTAGAAACATCTGGT----- 321
Db oArgGlySerProGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 733
QY 320 -----GCTGCCCAAAAGCTGCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 274
Db yProAlaGlyAlaAlaGlyGlnProGlyAlaLysGlyGlyGlyGlyGlyGlyGlyGly 753
QY 273 GACAGAACGACGCTGCTGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
Db sglyGlu-----AsnGlyileValGlyProThrGlySer----- 764
QY 213 TGGCTGTACATGTGGAAGATAGAGCCAGATGGCCATCTGGGAACACAGAGAGAGAG 154
Db -----ValGlyAlaAlaGlyProSerGlyProAlaGlyProGlyProValGly 781
QY 153 AGCTTCATTTTTTAGCAGTAGCATGTGCTGAAGTGGAAGACCTTGTGGCTTG---GG 97

Db 781 yserArgGly----- 793
QY 96 CTTCCCTGCTGCTTTTGAGAGG 75
Db 793 yPheProGlyAlaAlaGlyArg 800
RESULT 10
CAL12 HUMAN
ID CAL12 HUMAN STANDARD; PRT; 1418 AA.
AC P02458;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-NOV-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
GN COL2A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90067946; PubMed=2587267;
RA Su W.W., Lee B., Ramirez F., Machado M., Horton W.;
RT "Nucleotide sequence of the full length cDNA encoding for human type II procollagen.";
RL Nucleic Acids Res. 17:9473-9473(1989).
RN [2]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=87031574; PubMed=3021582;
RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
RT "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL Gene 44:11-16(1986).
RN [3]
RP SEQUENCE OF 432-1145 FROM N.A.
RA Ramirez F.;
RL Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 963-1418 FROM N.A.
RX MEDLINE=85190534; PubMed=3857598;
RA Cheah K.S.E., Stoker N.G., Griffin J.R., Grosfeld F.G., Solomon E.;
RT "Identification and characterization of the human type II collagen gene (COL2A1).";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN [5]
RP SEQUENCE OF 1120-1398 FROM N.A.
RX MEDLINE=85306861; PubMed=3840017;
RA Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J., Vuorio E.;
RT "Construction and identification of a cDNA clone for human type II procollagen mRNA.";
RL Biochem. J. 229:183-188(1985).
RN [6]
RP SEQUENCE OF 1106-1418 FROM N.A.
RX MEDLINE=88067771; PubMed=2825137;
RA Elima K., Vuorio T., Vuorio E.;
RT "Determination of the single polyadenylation site of the human pro alpha 1(II) collagen gene.";
RL Nucleic Acids Res. 15:9499-9504(1987).
RN [7]
RP SEQUENCE OF 1227-1289 FROM N.A.
RX MEDLINE=86104139; PubMed=3002437;
RA Nunez A.M., Francosano C., Young M.F., Martin G.R., Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for a human pro-alpha 1(II) collagen chain and demonstration of restriction fragment length polymorphism at the 3' end of the gene.";
RL Biochemistry 24:6343-6348(1985).
RN [8]
RP SEQUENCE OF 1176-1226 FROM N.A.
RX MEDLINE=84118798; PubMed=6320112;
RA Strom C.M., Upholt W.B.;
RT "Isolation and characterization of genomic clones corresponding to the human type II procollagen gene.";

RL Nucleic Acids Res. 12:1025-1038(1984).
 RN [9]
 RP SEQUENCE OF 35-167 FROM N.A.
 RX MEDLINE=89233138; PubMed=2714801;
 RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
 RT "Organization of the exons coding for pro alpha 1(I) collagen N-
 RT propeptide confirms a distinct evolutionary history of this domain N-
 RT the fibrillar collagen genes";
 RL Genomics 4:438-441(1989).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans";
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97259599; PubMed=9101280;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP VARIANT SER-1074.
 RX MEDLINE=90036309; PubMed=2572591;
 RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
 RA Hollister D.W.;
 RT "Glycine to serine substitution in the triple helical domain of pro-
 RT alpha 1 (II) collagen results in a lethal perinatal form of short-
 RT limbed dwarfism";
 RL J. Biol. Chem. 264:18265-18267(1989).
 RN [13]
 RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.
 RX MEDLINE=89266907; PubMed=2543071;
 RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
 RT "Identification of the molecular defect in a family with
 RT spondyloepiphyseal dysplasia";
 RL Science 244:978-980(1989).
 RN [14]
 RP VARIANT OSTEOARTHRTIS CVS-650.
 RX MEDLINE=90370826; PubMed=1975693;
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a
 RT cause of primary osteoarthritis associated with a mild
 RT chondrodysplasia";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
 RN [15]
 RP VARIANT OI-IV VAL-717.
 RX MEDLINE=91291136; PubMed=2064612;
 RA Bateman J.F., Hamagun M., Chan D., Cole W.G.;
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
 RT valine substitution in osteogenesis imperfecta type IV. Detection of
 RT the mutation and prenatal diagnosis by a chemical cleavage method.";
 RL Biochem. J. 276:765-770(1991).
 RN [16]
 RP VARIANT OSTEOARTHRTIS CVS-650.
 RX MEDLINE=91086471; PubMed=1985108;
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;
 RT "Cartilage expression of a type II collagen mutation in an inherited
 RT form of osteoarthritis associated with a mild chondrodysplasia";
 RL J. Clin. Invest. 87:357-361(1991).
 RN [17]
 RP VARIANT HYPOCHONDROGENESIS GUJ-984.
 RX MEDLINE=93054548; PubMed=1429602;
 RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
 RA Cohn D.H., Eyre D.R.;
 RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
 RT 1(I) chain produces hypochondrogenesis";
 RL J. Biol. Chem. 267:22522-22526(1992).
 RN [18]
 RP VARIANT HYPOCHONDROGENESIS SER-705.

RX MEDLINE=92262484; PubMed=1374906;
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
 RA Ramirez F., Vitale E., Lee B.;
 RT "Characterization of a type II collagen gene (COL2A1) mutation
 RT identified in cultured chondrocytes from human hypochondrogenesis";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
 RN [19]
 RP VARIANT WS-II ASP-198.
 RX MEDLINE=93304428; PubMed=8317498;
 RA Koerkoe J., Ritvaniemi P., Haataja L., Kaeeriaeinen H.,
 RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
 RT for glycine alpha 1-67 and that causes cataracts and retinal
 RT detachment: evidence for molecular heterogeneity in the Wagner
 RT syndrome and the Stickler syndrome (arthro-ophthalmopathy)";
 RL Am. J. Hum. Genet. 53:55-61(1993).
 RN [20]
 RP VARIANT SEMD CYS-840.
 RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,
 RA Eyre D.R.;
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces
 RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
 RL Am. J. Hum. Genet. 53:A209-A209(1993).
 RN [21]
 RP VARIANT OSTEOARTHRTIS CVS-650.
 RX MEDLINE=93282819; PubMed=8507190;
 RA Holderbaum D., Mailemud C.J., Moskowitz R.W., Haqqi T.M.;
 RT "Human cartilage from late stage familial osteoarthritis transcribes
 RT type II collagen mRNA encoding a cysteine in position 519";
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
 RN [22]
 RP VARIANT SEMD ASG-285.
 RX MEDLINE=93252400; PubMed=8486375;
 RA Vikkula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
 RA Peltonen L.;
 RT "A mutation in the amino-terminal end of the triple helix of type II
 RT collagen causing severe osteochondrodysplasia";
 RL Genomics 16:282-285(1993).
 RN [23]
 RP VARIANT SEDC CYS-206.
 RX MEDLINE=94063862; PubMed=8244341;
 RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
 RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
 RT family with an Arg75-->Cys mutation in the procollagen type II gene
 RT (COL2A1)";
 RL Hum. Genet. 92:499-505(1993).
 RN [24]
 RP VARIANT SEDC CYS-920.
 RX MEDLINE=93315508; PubMed=8325895;
 RA Chan D., Taylor T.K.F., Cole W.G.;
 RT "Characterization of an arginine 789 to cysteine substitution in
 RT alpha 1(II) collagen chains of a patient with spondyloepiphyseal
 RT dysplasia";
 RL J. Biol. Chem. 268:15238-15245(1993).
 RN [25]
 RP VARIANT SEDC SER-1128.
 RX MEDLINE=93140139; PubMed=8423604;
 RA Cole W.G., Hall R.K., Rogers J.G.;
 RT "The clinical features of spondyloepiphyseal dysplasia congenita
 RT resulting from the substitution of glycine 997 by serine in the alpha
 RT 1(II) chain of type II collagen";
 RL J. Med. Genet. 30:27-35(1993).
 RN [26]
 RP Alignment Scores:
 Pred. No.: 0.0282 Length: 1418
 Score: 138.50 Matches: 87
 Percent Similarity: 35.47% Conservative: 18
 Best Local Similarity: 29.39% Mismatches: 119
 Query Match: 3.74% Indels: 72
 DB: 1 Gaps: 17
 US-09-645-078-1 (1-2043) x CA12_HUMAN (1-1418)


```

RP SEQUENCE OF 529-567 FROM N.A.
RX MEDLINE=85122694; PubMed=6395893;
RA Genovese C., Rowe D., Kream B.;
RT "Construction of DNA sequences complementary to rat alpha 1 and alpha
RT 2 collagen mRNA and their use in studying the regulation of type I
RT collagen synthesis by 1,25-dihydroxyvitamin D.";
RL Biochemistry 23:6210-6216(1984).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PTM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains.
CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC the oxygen atom of a post-translationally added hydroxyl group.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11432; AAA0832.1; ALT SEQ.
CC InterPro; IPR008161; C1g_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR001007; WVF_C.
CC ProDom; PD000007; C1g_helix; 1.
CC PROSITE; PS01208; WVF_C_1; PARTIAL.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Pyridinolone carboxylic acid (PROBABLE).
FT MOD_RES 1 1 PYRIDINONE CARBOXYLIC ACID (PROBABLE).
FT MOD_RES 9 9 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT INVOLVED IN CROSS-LINKING.
FT MOD_RES 28 28 HYDROXYLATION (PROBABLE).
FT MOD_RES 31 31 HYDROXYLATION (PROBABLE).
FT MOD_RES 34 34 HYDROXYLATION (PROBABLE).
FT MOD_RES 43 43 HYDROXYLATION (PROBABLE).
FT MOD_RES 46 46 HYDROXYLATION (PROBABLE).
FT MOD_RES 49 49 HYDROXYLATION (PROBABLE).
FT MOD_RES 103 103 HYDROXYLATION.
FT CARBOHYD 103 103 O-LINKED (GAL. .).
FT MOD_RES 424 424 HYDROXYLATION (PROBABLE).
FT MOD_RES 547 547 HYDROXYLATION (PROBABLE).
FT NON_CONS 567 568
FT DOMAIN 641 651 MAJOR ANTIGENIC DETERMINANT (OF NEUTRAL
FT SALT-EXTRACTED RAT SKIN COLLAGEN).
SQ SEQUENCE 671 AA; 60615 MW; 9DC3114204AC4918 CRC64;

Alignment Scores:
Pred. No.: 0.0245 Length: 671
Score: 138.00 Matches: 101
Percent Similarity: 28.50% Conservative: 19
Best Local Similarity: 23.99% Mismatches: 137
Query Match: 3.73% Indels: 165
DB: 1 Gaps: 23

US-09-645-078-1 (1-2043) x CALL_RAT (1-671)
QY 1144 GGCAGAACCCAGCGCCAGCCTGGAGACATTAAGGCATCCCTGGCAT----- 1095
Dd 80 GlyGlnArgGlyProGlyProGlyProGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeu 99
QY 1094 -----TTGTGTGGAAGCGTGGTCAACCCATGC 1068
Dd 100 ProGlyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAsnThrGly 119
QY 1067 CCTTGCCCTGGGTGATGTTATGACCCAGGTCTGAAGATGGGCA-----AGA 1020

```

```

Dd 120 ProAla-----GlyProLysGlyGluProGlySerProGlyGluAsnGly 134
QY 1019 ATTCCAATCCACGAATTATCATCTCGGGAAGTCTGGCCACACAGGGCTC----- 969
Dd 135 AlaPro-----GlyGlnMetGlyProArgGlyLeuProGlyGlu 147
QY 968 -----GAGCCAGGTCTCT-----CATAGCCACACAAGCA 942
Dd 148 ArgGlyArgProGlyProGlySerAlaGlyAlaArgGlyAspGlyAlaValGly 167
QY 941 GGTAGCGTTCTCTGAGGGCTTGGGCAAGACTGAGTGTCTTGTAGATCTCCAGCTGGC 882
Dd 168 AlaAlaGlyProProGlyProThrGlyProThrGly-----ProProGly 182
QY 881 TTGGCAGATGACCTGCATCATAGTAGGGTTGGTCTCT----- 843
Dd 183 PheProGlyAlaAlaGlyAlaLysGlyGluAlaGlyProGlnGlyAlaArgGlySerGlu 202
QY 842 ---CCTTCTTGTAGTTTGTCTCATGCTGCCCATCATCAATGCGACTGTCAATCATGAGAT 786
Dd 203 GlyProGlnGlyValArgGlyGluProGlyProGly----- 215
QY 785 CTCCCTTTGTGGCTTCTCGGGAACGAAACACGGCCGGGGTCCCGA---CCAGGTGCA 729
Dd 216 ---ProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnProGlyAla 234
QY 728 CGATATGCGAGTTGAGGGGCTCTTTCAGCAGCG-----GCTAGAGGG 684
Dd 235 LysGlyAlaAsnGlyAlaProGlyLleAlaGlyAlaProGlyPheProGlyAlaArgGly 254
QY 683 ACTGAGGTGAAGAAGCGCACCTCTTGTAGACACACAGTGGTGT----- 639
Dd 255 ProSerGly-----ProGlnGlyProSerGlyAlaProGlyProLysGly 269
QY 638 AGGAGCGGCGAGCTTCTCCACACCTCAAGGGCTGTGACTGCACAGAGCGCTCAGT 579
Dd 270 AsnSerGlyGluProGlyAlaProGlyAsnLysGly-----AspThrGlyAlaLysGly 287
QY 578 GAGCCCGGGGATGATTTTCATCTTGTGGATGATGTCACAGGAGTGCAGAACACAGGG 519
Dd 288 GluProGly-----ProAlaGlyValGln-----Gly 236
QY 518 CCGGCTGTCTCCACTGAAAGAGCTGAGTGTCTCCGGGACACAGGTTCATGTAGG 459
Dd 297 ProProGlyProAlaGlyGluGluGly----- 305
QY 458 CATCAAGAGCGCTCATGTCACCAAGAAGACGGCCCGTATCATAGTCCGACAGCATGT 399
Dd 306 -----LysArgGlyAlaArgGlyGluProGlyProSerGlyLeuProPro----- 322
QY 398 GCAGCATCCAGGGGTCTCTGCTTGNAGTCTCCACAGCTGCCAGGGGGCTCCATCA 339
Dd 323 -----GlyGluArgGlyGlyProGly 329
QY 338 GGTAGAAAACATCTGGGTGCTGCCCAAAAGCTGCCCAAAAGAGAGCCAGAGGCC 279
Dd 330 SerArgGlyPheProGlyAlaAspGlyValAlaGlyPro-----LysGlyProAlaGly 347
QY 278 AGGAAGACAGAACCCAGCACGTGATGCGCTCGGGCTGTGCTTCATAGACAGGAGCTGA 219
Dd 348 GluArgGlySerProGly----- 353
QY 218 TGTGTGCTGTACATGTGGAAGAAATAGAGCCAGATGGCCATCTGGGAAACACAGAAACA 159
Dd 354 -----ProAlaGlyProLysGlySerProGlyGlu 363
QY 158 GCAGAGCTTCATTTTATAGGCGATGATGCTGAAGTGGAGACCTTGTGGCTTG 99
Dd 364 Ala-----GlyArgProGlyGluAla 370
QY 98 GGTCTCCCTGCTGCTTTGTAGAGGCTCACTGCTAGCCGGGCCATCGCTTTTCTCTCTTA 39
Dd 371 GlyLeuProGlyAla---LysGlyLeuThrGlySerProGly-----SerProGly 386

```

```

QY      38 CCC 36
Db      387 PRO 387

RESULT 12
CALL CHICK
ID      CALL CHICK STANDARD; PRT; 1453 AA.
AC      P02457;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Collagen alpha 1(I) chain precursor.
GN      COL1A1.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OC      NCBI_TaxID=9031;
OX      [1]
RN      SEQUENCE OF 1-153 FROM N.A.
RP      MEDLINE=88056316; PubMed=3678834;
RX      Finer M.H., Boedtker H., Doty P.;
RA      "Construction and characterization of cDNA clones encoding the 5' end
RT      of the chicken pro alpha 1(I) collagen mRNA.";
RL      Gene 56:71-78(1987).
RN      [2]
RP      SEQUENCE OF 1-144 FROM N.A.
RX      MEDLINE=88007542; PubMed=2820966;
RA      Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RX      Kang A.H., Gross J.;
RA      "Unusual DNA sequences located within the promoter region and the
RT      first intron of the chicken pro-alpha 1(I) collagen gene.";
RL      J. Biol. Chem. 262:13323-13332(1987).
RN      [3]
RP      SEQUENCE OF 152-1187.
RX      MEDLINE=82231995; PubMed=7093229;
RA      Highberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RX      Kang A.H., Gross J.;
RA      "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
RT      complete primary structure of the helical portion of the chick skin
RL      collagen alpha 1(I) chain.";
RL      Biochemistry 21:2048-2055(1982).
RN      [4]
RP      SEQUENCE OF 1200-1205.
RX      MEDLINE=72243016; PubMed=5047697;
RA      Eyre D.R., Glimcher M.J.;
RT      "Evidence for a previously undetected sequence at the carboxyterminus
RL      of the alpha 1 chain of chicken bone collagen.";
RL      Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN      [5]
RP      SEQUENCE OF 981-1453 FROM N.A.
RX      MEDLINE=81160715; PubMed=6927845;
RA      Fuller F., Boedtker H.;
RT      "Sequence determination and analysis of the 3' region of chicken pro-
RT      alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT      including the carboxy-terminal propeptide sequences.";
RL      Biochemistry 20:996-1006(1981).
RN      [6]
RP      SEQUENCE OF 1311-1453 FROM N.A.
RX      MEDLINE=80134546; PubMed=6987088;
RA      Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA      Pastan A., Decrombrugne B., Fietzek P.P., Olsen B.R.;
RT      "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT      carboxyl end of pro alpha 1(I)-chains.";
RL      FEBS Lett. 111:61-65(1980).
RN      [7]
RP      FUNCTION: Type I collagen is a member of group I collagen
CC      (fibrillar forming collagen).
CC      -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC      -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC      bones. In bones the fibrils are mineralized with calcium
CC      hydroxyapatite.
CC      -1- PTM: Proline residues at the third position of the tripeptide
CC      repeating unit (G-X-Y) are hydroxylated in some or all of the

```

```

CC      chains. Pro-1153 is the only 3-hydroxypro and the only
CC      hydroxylated proline in position X.
CC      -1- SIMILARITY: Contains 1 VWF domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M17839; AAA48704.1; -.
CC      EMBL; M17838; AAA48704.1; JOINED.
CC      EMBL; V00401; CAA23695.1; -.
CC      EMBL; M10571; AAA48671.1; ALT_SEQ.
CC      EMBL; M17607; AAA48672.1; -.
CC      PIR; A21179; A27179.
CC      PIR; I50629; I50629.
CC      InterPro; IPR008161; Clg_helix.
CC      InterPro; IPR008160; Collagen.
CC      InterPro; IPR000885; Fib_collagen_C.
CC      InterPro; IPR002181; Fibrinogen_C.
CC      InterPro; IPR001007; VWF_C.
CC      Pfam; PF01410; COLFI; 1.
CC      Pfam; PF01391; Collagen; 18.
CC      Pfam; PF00093; vwc; 1.
CC      ProDom; PD000007; Clg_helix; 2.
CC      ProDom; PD002078; Fib_collagen_C; 1.
CC      SMART; SM00038; COLFI; 1.
CC      SMART; SM00214; VWC; 1.
CC      PROSITE; PS01208; VWF_1; 1.
CC      PROSITE; PS50184; VWF_2; 1.
CC      KW      Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC      Glycoprotein; Collagen; Signal; Pyrrolidone carboxylic acid.
CC      FT      SIGNAL 1 22
CC      FT      PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
CC      FT      CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.
CC      FT      PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
CC      FT      DOMAIN 31 89 VWF
CC      FT      MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
CC      FT      MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
CC      FT      MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
CC      FT      MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
CC      FT      MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
CC      FT      MOD_RES 1153 1153 HYDROXYLATION.
CC      FT      CONFLICT 1187 1187 F -> L (IN REF. 5).
CC      FT      CONFLICT 1441 1441 Q -> H (IN REF. 6).
CC      SQ      SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Alignment Scores:
Pred. No.: 0.0309 Length: 1453
Score: 138.00 Matches: 112
Percent Similarity: 31.31% Conservative: 22
Best Local Similarity: 26.17% Mismatches: 166
Query Match: 3.73% Indels: 129
DB: 1 Gaps: 25

US-09-645-078-1 (1-2043) x CALL_CHICK (1-1453)
QY      1144 GGCAGAACAGCCAGCCGCTGGAGACATTAAAGGCATCCTGGCAT-----TTG 1092
Db      114 GlyAspArgGlyLeuProGlyProProGlyArgAspGlyLeuProGlyGlnProGlyLeu 133
QY      1091 TGTGAAAGCGTGTCCACCA---TGCCCTGCCTCGGTGATGTTATGCACCCAGG--- 1038
Db      134 ProGlyProProGlyProProGlyProProGlyLeuGlycylAsnPheAlaProGlnMet 153
QY      1037 -----TCTGAAGATGGGCAAGATCCCAATCCCAAGCAATTCATACATTCGGG 990
Db      154 SerTyrGlyTyrAspGluLysSerAlaGlyValAlaValPro-----GlyProMetGly 171
QY      989 AAGTCTGGCCACAGGGGGCTC-----GAGCCAGGTCTCTCATAGCCACAA 945

```

Db QY 172 ProAlaGlyProArgGlyLeuProGlyProGlyAlaProGlyProGlnGlyPheGln 191
:::|||||:::|||||
Db QY 944 -----CGAGTACGTTCTCTGCAGGCGCTTGGCGAAG 912
|||||
Db 192 GlyProProGlyGluProGlyGluProGlyAlaSerGlyProMetGlyPro 208
|||||
Db QY 911 ACTGATGCTCTGTAGATCTCCAGCTGC 882
|||||
Db 209 -----ArgGlyProAlaGlyProProGlyLysAsnGlyAspGlyGlu 223
|||||
Db QY 881 TTGGCAGATGACCTGCATCATAGTAGGTTGCTCTCTCTTCTGAGTTTTCCTCAT 822
|||||
Db 224 AlaGlyLys-----ProGlyArgProGlyGlnArgGlyProProGlyProGlnGlyAlaArg 242
:::|||||
Db QY 821 GCTGCCCATCAATGCGACTGTCAATCATGAGATCTCCCT 780
|||||
Db 243 GlyLeuProGlyThrAla-----GlyLeuProGlyMetLysGlyHisArg 257
|||||
Db QY 779 -----TTGTGGTCTCTGGGACCGAACACAGCGCCGGGGTCCCGACACAGGTGCAGA 726
:::|||||
Db 258 GlyPheSerGlyLeuAspGlyAlaLysGlyGlnProGlyProAlaGlyProLysGlyGlu 277
|||||
Db QY 725 TATGAGGTTGAGGAGGGGCTTTTCAGCAGCGGGT-----AGAGGAGCTGCA 678
|||||
Db 278 ProGlySerProGlyGluAsn-----GlyAlaProGlyGlnMetGlyProArgGlyLeuPro 296
|||||
Db QY 677 GGTGAAGAGCGCACT---CCTTGACACACAGTGGCTGTAGAGCGGCGAGCGCTTCT 621
|||||
Db 297 GlyGluArgGlyArgProGlyProSerGlyProAlaGlyAlaArgGlyAsnAspGlyAla 316
|||||
Db QY 620 CCACCACCTCAAAGGGCTGTACTGCACAGGAGCTGCAGTGAGCCCGGGGATGATT 561
|||||
Db 317 Pro-----GlyAlaAlaGlyProProGly----- 324
|||||
Db QY 560 CATCTTGTGGATGATGTCACAGGAGTGCAGACACAGGCGCGGTCTTCTCCCACT 501
|||||
Db 325 -----ProThrGlyProAlaGlyProProGly 333
|||||
Db QY 500 GAAGAGGCTGGACT---GTCTCCGGGACCGAGTTCCA-----TGTAGCATCAAGA 450
|||||
Db 334 PheProGlyAlaAlaGlyAlaLysGlyGluThrGlyProGlnGlyAlaArgGlySerGlu 353
|||||
Db QY 449 CGCTCATGTGCGACAGAGAGCGGCC-----GTATCATGATCCC 411
:::|||||
Db 354 GlyProGlnGlySerArgGlyGluProGlyProProGlyProAlaGlyAlaAlaGlyPro 373
|||||
Db QY 410 GCACAGCATGTGCGATCCAGCGGTCTCTGCTGAGTCACTCCACAGCTGCC--- 354
|||||
Db 374 AlaGlyAsnProGlyAlaAspGlyGlnProGlyAla---LysGlyAlaThrGlyAlaPro 392
|||||
Db QY 353 -----AGCGGGCTCCATCAGGTAGAAAACATCTG 324
|||||
Db 393 GlyIleAlaGlyAlaProGlyPheProGlyAlaArgGlyProSerGlyProGlnGlyPro 412
|||||
Db QY 323 GTGTCTGCC-----CAAAAGCTGCCCCCAAAAGAGAGC 288
|||||
Db 413 SerGlyAlaProGlyProLysGlyAsnSerGlyGluProGlyAlaProGlyAsnLysGly 432
|||||
Db QY 287 CAGAGCGCCAGGAACAGACAGCAGCTGCATCGCTCGCGGTGCTCCCTCATAGACA 228
|||||
Db 433 AspThrGlyAlaLysGlyGluProGlyProAlaGlyValGlnGlyProProGlyProAla 452
|||||
Db QY 227 GGGAGCTGATGTGTGCTGTACATGTGAAGAATAGACCGAAGATGCCATCTGGGAAA 168
|||||
Db 453 GlyGlu-----GluGlyLysArgGlyAlaArgGlyGluPro---GlyProAlaGlyLeu 469
|||||
Db QY 167 CAGAAACAGCAGGAGCTTCATTTTTTGGCAGTAGCATTTGCTGAAGTGGAGAGCT 108
|||||
Db 470 ProGlyProAlaGlyGlu-----ArgGlyAlaPro 479
|||||
Db QY 107 TGTGGCTTGGCTTCCCTCTCTCT 84
|||||

Db 480 GlySerArgGlyPheProGlyAla 487
RESULT 13
MUSA HUMAN
ID P38088; O76065; Q13792; Q8WQ03; PRT; 1233 AA.
AC 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 5AC (Mucin 5 subtype AC, tracheobronchial) (Tracheobronchial
mucin) (TBM) (Major airway glycoprotein) (Fragment).
GN MUC5AC OR MUC5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE OF 1-372 FROM N.A.
RX MEDLINE=21426417; PubMed=1153137;
RA Escande P., Aubert J.-P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
repetitive region.";
RL Biochem. J. 358:763-772 (2001).
RN [2]
RP SEQUENCE OF 193-1233 FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Lung;
MEDLINE=94230376; PubMed=7513696;
RA Meerraman D., Charles P., Daskal E., Polymeropoulos M.H.,
Martin B.M., Rose M.C.;
RT "Cloning and analysis of cDNA encoding a major airway glycoprotein,
human tracheobronchial mucin (MUC5).";
RL J. Biol. Chem. 269:12932-12939 (1994).
RN [3]
RP SEQUENCE OF 153-1233 FROM N.A.
RX TISSUE=Placenta, and Trachea;
MEDLINE=98285541; PubMed=9620876;
RA Buisine M.P., Desseyn J.L., Porchet N., Degand P., Laine A.,
Aubert J.-P.;
RT "Genomic organization of the 3'-region of the human MUC5AC mucin gene:
additional evidence for a common ancestral gene for the 1p15.5 mucin
gene family.";
RL Biochem. J. 332:729-738 (1998).
RN [4]
RP SEQUENCE OF 284-1233 FROM N.A.
RX MEDLINE=95293957; PubMed=7775418;
RA Lesuffleur T., Roche F., Hill A.S., Lacasa M., Fox M., Swallow D.M.,
Zweibaum A., Real F.X.;
RT "Characterisation of a mucin cDNA clone isolated from HT-29 mucus
secreting cells: The 3' end of MUC5AC?";
RL J. Biol. Chem. 270:13665-13673 (1995).
CC - FUNCTION: MAJOR AIRWAY GLYCOPROTEIN.
CC - SUBUNIT: Multimeric.
CC - SUBCELLULAR LOCATION: Secreted.
CC - TISSUE SPECIFICITY: TRACHEOBRONCHIAL MUCOSAE (NOT SOLELY).
CC - SIMILARITY: TO VARIOUS MUCINS.
CC - SIMILARITY: Contains 1 WFPC domain.
CC - SIMILARITY: Contains 1 WFCD domain.
CC - SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
EMBL; AJ298319; CAC83676.1; -
EMBL; U06711; AAA18431.1; -
EMBL; AJ001402; CAA04737.1; -
EMBL; AJ001403; CAA04738.1; -
EMBL; Z48314; CAA88307.1; -
Genew; HGNC:7515; MUC5AC.

DR GO: GO:0005201; F:extracellular matrix structural constituent; TAS.
DR GO: GO:0007155; P:cell adhesion; NAS.
DR InterPro: IPR006208; Cys_knot.
DR InterPro: IPR006207; Cys_knot_C.
DR InterPro: IPR002919; TIL_Cysrich.
DR InterPro: IPR001007; VWF_C.
DR InterPro: IPR001846; VWF_D.
DR Pfam: PF00007; Cys_knot; 1.
DR Pfam: PF00093; vwf; 1.
DR Pfam: PF00094; vwd; 1.
DR PROSITE: PS01185; CTCK_1; 1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS01208; VWF_1; 2.
DR PROSITE: PS0184; VWF_2; 2.
KW Repeat; Glycoprotein.
FT NON_TER 1
FT DOMAIN 431 496 VWF 1.
FT DOMAIN 500 659 VWF.
FT DOMAIN 688 712 5 X 5 AA TANDEM REPEATS OF T-T-V-G-P/S.
FT REPEAT 688 692 1.
FT REPEAT 693 697 2.
FT REPEAT 698 702 3.
FT REPEAT 703 707 4.
FT REPEAT 708 712 5.
FT DOMAIN 857 923 VWF 2.
FT DOMAIN 962 1026 VWF 3.
FT DOMAIN 1111 1199 CTCK.
FT DISULFID 1111 1161 BY SIMILARITY.
FT DISULFID 1136 1175 BY SIMILARITY.
FT DISULFID 1140 1191 BY SIMILARITY.
FT DISULFID 1158 1193 BY SIMILARITY.
FT DISULFID 1160 1198 BY SIMILARITY.
FT CONFLICT 193 194 VS -> HE (IN REF. 2).
FT CONFLICT 406 406 P -> R (IN REF. 4).
FT CONFLICT 463 465 SER -> REP (IN REF. 2).
FT CONFLICT 478 478 G -> A (IN REF. 2).
FT CONFLICT 492 494 HHYQCVCVSGMDHYHFDQTYTFDLNCTYVLVQGI
FT VYGHFVLVDNLYFCGAEDGLSPRSIIILEYHODRVLLTRK
FT VHGVTNIIENKVVSECFR -> IITTSASVCAAGVPT
FT TSPSTAPTPSTWTTARTLGAADCARVWPLPRARRQLLRG
FT GRALLPEVHHGVPVPGRGADPQASPRGVGDKRDLHQGGQ
FT PLPL (IN REF. 2).
FT G -> A (IN REF. 4).
FT H -> Q (IN REF. 3).
FT A -> R (IN REF. 2).
FT GIMCFPTPADRVYQFCG -> RTCAPHLPSRQGVPLR
(IN REF. 2).
FT R -> P (IN REF. 4).
FT S -> T (IN REF. 3).
FT G -> R (IN REF. 2).
FT A -> R (IN REF. 2).
FT G -> R (IN REF. 2).
FT SQCCGTCVQACVINTSKSPAHLPYQGE -> RRSVAVP
FT CRSPVSPPARAPPTSTLAS (IN REF. 2).
FT L -> P (IN REF. 2).
FT C -> S (IN REF. 2).
FT GTOHSEAEPEPSQEAESGWGVPVSPMH -> ATPST
FT RRRRPSPARRVRVAGREASVPHALS (IN REF.
2).
FT VPSPMH -> QCPPTDQHRPPDLQGPPIPLSSASKAS
FT GTCAPVQAAALNTLSFAFLWRVWANGHLPLGGGALTHPA
FT CSHLSGAPGLAEILLWPCIQPAVLGT (IN REF. 4).
FT CS SEQUENCE 1233 AA; 130072 MW; 37CEA5FADCA828DE CRC64;
Alignment Scores:
Pred. No.: 0.032 Length: 1233
Score: 137.50 Matches: 123
Percent Similarity: 32.76% Conservative: 48
Best Local Similarity: 23.56% Mismatches: 197
Query Match: 3.67% Indels: 154
DB: 1 Gaps: 24
US-09-645-078-1 (1-2043) x MUSA_HUMAN (1-1233)

QY 426 CCGTCTTCTGTGCGACATGAGCGTCTTGTAGTGCCTACATGGAACCTG-----GTCCCC 479
DB 29 ProSerGlyProGlyThrThrProSerProValProThrThrSerThrThrSerAlaPro 48
QY 480 GGAGACAGTCCAGCCTCTTTCAGTGGGAGAAACAGCCGGCCCTGTGTCTGCACCTGCCT 539
DB 49 ThrThrSerThrThrSerGlyProGlyThrThrProSerPro-----ValPro 64
QY 540 GTGACATCA-----TCCCAACAAGTGAATCATCCCCGGGCTC 578
DB 65 ThrThrSerThrThrSerAlaProThrThrThrThrThrSerAlaSerThrAlaSerThr 84
QY 579 ACTGCAGGCTCTGTGTCAGTCAACAGCCCTTTAGGTGGTGGAGAGAGCCCTGCCCTCTCT 638
DB 85 ThrSerGlyProGlyThrThrProSerProValProThrThrThrThrThrSerAlaPro 104
QY 639 ACAGCCACGCTGTGCTCAAGAGGTGCGCTTCTCACTCACTGAGTCCCTCTACCCGCTGC 698
DB 105 ThrThrSerThrThrSerAlaSerThrAlaSerThrThrSerGlyProGlyThrSerLeu 124
QY 699 TGAAGACCCCT---CCCTCAACCTGCATATCGTGCACCTGTGCGGACCCCGGCGCG 755
DB 125 SerProValProThrThrSerThrThrSerAlaProThrThrThrThrThrSerGlyPro 144
QY 756 TGTTCCTGTTCCGAGAACCCACAAAGGAGATCTCATGATTGACAGTGCATGTGATGG 815
DB 145 GlyThrThrProSerProVal----- 151
QY 816 GGCAGCATGACAAACAACTCAAGAGGAGGACCAACCTACTATGTGATGCAAGTCACT 875
DB 152 -----ProThrThrSerThrThrSerAla 159
QY 876 GCCAAGCCAGCTGGAGATCTCAAGACATCCATCCATCTGTCGCCAAGGCCCTGCAAGAAC 935
DB 160 ProThrThrSerThrThrSerGlyProGlyThrThrProSerProValProThrThrSer 179
QY 936 GCTACC----- 941
DB 180 ThrThrProValSerLysThrSerThrSerHisLeuSerValSerLysThrThrHisSer 199
QY 942 -----TGC-----TTGTGCGTATGAGGACCTGTGCTCGAGCCCTCG 977
DB 200 GlnProValThrSerAspCysHisProLeuCysAlaTyrThrLysTyrPheAspValAsp 219
QY 978 TGCCCCAGACTT---CCGAATGTATGAATTCGTGGATGGAATCT----- 1022
DB 220 PheProSerProGlyProHisGlyAspLysGluThrTyrAsnAsnIleIleArgSer 239
QY 1023 -----TGCCCATCTTCAGACCTTCAGCTGCGTAAACATCACCCGAGCAAGGCA 1070
DB 240 GlyGluLysIleCysArg-----ArgProGluGluIleThrArgLeuGlnCysArgAla 257
QY 1071 TGGGTG---ACCAGCTTTCACACAAATGCCAGGAGATCCCTTAATGTCTCCAGGCTT 1127
DB 258 GluSerHisProGluValAsnIleGluHisLeuGlyGlnValGlnValGlnCysSerArgGlu 277
QY 1128 GGCCTGCTTTCG-----CCTATGAAAAGGTTTCTCGACTTC 1166
DB 278 GluGlyLeuValCysArgAsnGlnAspGlnGlnGlyProPheLysMetCysLeuAsnTyr 297
QY 1167 AGAAGCCTGTGCGATGCCATGAATTCGTGGCTACCCGACGCTCAGATCTGAACAAG 1226
DB 298 GluValArgVal-----LeuCysGluThr-ProArgGlyCys----- 310
QY 1227 AACAGAGAACCTGTGCTGGATCTCTGTCTACTGCTGACTGCTCCCTGAGCAATCCACT 1286
DB 311 -----ProValThrSerValThrProTyrGlyThrSerPro---ThrAsnAlaLeu 326
QY 1287 AAGAGGGTTGAGAAGGCTTGTGTCACCTGTGTGTCAGCTCAGTCACTTCTCTGAATG 1346
DB 326 u-----TyrProSerLeuSerThrSerMetValSerAlaSerVal----- 339


```

DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 19.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; Clg_helix; 5.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI_1.
DR SMART; SM00214; VWC_1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
KW Disease mutation; 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 1226
FT PROPEP 1227 1496
FT DOMAIN 39 97
FT MOD_RES 230 290
FT MOD_RES 293 293
FT MOD_RES 296 296
FT MOD_RES 608 608
FT MOD_RES 614 614
FT VARIANT 960 960
FT CONFLICT 292 292
FT CONFLICT 1418 1418
FT CONFLICT 1438 1438
FT CONFLICT 1460 1460
FT CONFLICT 1496 1496
FT SEQUENCE 1496 AA; 144720 MW; 82827C17A86644F5A CRC64;

Alignment Scores:
Pred. No.: 0.034 Length: 1496
Score: 137.50 Matches: 88
Percent Similarity: 33.44% Conservative: 12
Best Local Similarity: 29.43% Mismatches: 107
Query Match: 3.72% Indels: 92
DB: 1 Gaps: 21

US-09-645-078-1 (1-2043) x CA25_HUMAN (1-1496)
QY 1136 ACCAGCGCAAGCTGGGAGACATTAAAGGCATCTCGCATTTGTGGAAAGCGTGGT 1077
DQ 439 ThrSerGlyProGlySer-----AlaGlyProGly 450
QY 1076 CACCCATGCCCT-----TGCTCGGTGATGTTATGACCCAGCTCTGAAGATGGGCA 1023
DQ 451 SerProGlyProGlySerThrGly-----ProGlyAsnSerGly--- 465
QY 1022 AGAATTCGAATCCCAAGATTTCATATC-----GGGAAGTCTGGCCACAGGGCTCGAG 966
DQ 466 -----LeuProGlyAspProGlyPheLysGlyGluAlaGlyProLysGly---Glu 481
QY 965 CCAGGTCCTCATAGCCACAGCAGGATAGCTTCTCGAGGCCCTTGGCAGAGGACTGGA 906
DQ 482 ProGlyProHisGlyIleGly-----ProIleGlyProGlyGluGly 498
QY 905 TGCTCTGTAGATCTCCAGCTGCTTGGCAGATGACCTGCATCATAGTAGTGGTGGT 846
DQ 499 -----LysArgGlyProArgGly-----AspProGlyThrLeuGly 510
QY 845 CTTCTCTTGTAGTTTGTCTCATGCTGCCCATCAATGCGACTGTCAATCATGAGAT 786
DQ 511 ProProGlyProValGlyGluArgGlyAlaProGlyAsnArgGlyPheProGlySerAsp 530
QY 785 -----CTCCCT 780
DQ 531 GlyLeuProGlyProLysGlyAlaGlyGluArgGlyProValGlySerGlyPro 550
QY 779 TTGTGCTTCTCGGAAACGGAACACAGCGCGGGGTCCCGGA---CCAGGTGCACCATAT 723

```

```

DQ 551 LysGlySerGlnGlyAspProGlyArgProGlyGluProGlyLeuProGlyAla----- 568
QY 722 GCAGGTTGAGGAGGGGTCTTTTCAGCA-----GCGGGTAGAGGAGGACTGCA 678
DQ 569 -----ArgGlyLeuThrGlyAsnProGlyValGlnGlyProGlyLysLeu 584
QY 677 GGTGAAGAAGCGCACCTCTTGGACACCGAGCTGTAGAGCGGCGGAGGCTCTTCCCA 618
DQ 585 GlyProLeuGlyAlaProGly-----GluAspGlyArgProGlyPro 598
QY 617 CCACCTCAAGAGGCTGTGACTGCACAGGAGCGCTGCAGTGCAGCGGGGATGATTTTCAT 558
DQ 599 Pro-----GlySerIleGlyLeuGlyGlnProGly----- 609
QY 557 CTTGTGGGATGATGCACAGGAGGTGCAGAACAGAGCGGCGGCTGTCTCCCACTGAA 498
DQ 610 ThrMetGlyLeuProGlyProLysGlySerAsnGlyAspProGly---LysAspGlyGlu 628
QY 497 AGAGGCTGGACT---GTCTCGGAGGAGGAGGCTTCATGATGAGCATCAAGAGCGCTCATGT 441
DQ 629 AlaGlyAsnProGlyVal-ProGlyGlnArgGlyAlaProGly---LysAspGlyLysVa 647
QY 440 CGCACAAAGAGAGCGGCGGTATCATGATCCCG-----CACAG 405
DQ 647 I-----GlyProTyrGlyProGlyProGlyProGlyLeuArgGlyGluArgGly 663
QY 404 CATGTGACGATCAGCGGCGGTCTCTGTTCAAGTGCATCCACAGCGGCCAGGC 350
DQ 663 yGluGlnGlyProGlyProGlyProThrGlyPheGlnGlyHisProGlyProGly 681

RESULT 15
CA21_CANFA
ID CA21_CANFA STANDARD; PRT; 1366 AA.
AC 046352;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF035120; AAC64485.1;
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 17.
CC ProDom; PD000007; Clg_helix; 3.

```

ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SMO038; COLFI; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Collagen; Signal; Pyrrolidone carboxylic acid.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE (BY
SIMILARITY).
FT CHAIN 80 1102 COLLAGEN ALPHA 2(I) CHAIN.
FT PROPEP 1103 1366 CARBOXYL-TERMINAL PROPEPTIDE (BY
SIMILARITY).
FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID (BY
SIMILARITY).
FT MOD_RES 84 84 CONVERTED TO AN ALDEHYDE GROUP THAT IS
INVOLVED IN CROSS-LINKING (BY
SIMILARITY).
SQ SEQUENCE 1366 AA; 129400 MW; CD936969E08BCD4 CRC64;

Alignment Scores:
Pred. No.: 0.036 Length: 1366
Score: 137.00 Matches: 113
Percent Similarity: 30.14% Conservative: 35
Best Local Similarity: 23.01% Mismatches: 191
Query Match: 3.70% Indels: 152
DB: 1 Gaps: 20

US-09-645-078-1 (1-2043) x CA21_CANFA (1-1366)

QY	1180	GCCACAGCCTTCTGAAGTCGAGAACTTTTCATAGGGCAAGACGACGCCAAGCGTG	1121
DV			
Db	335	AlaAlaGlyAlaThrGlyAlaArgylleValGlyGluProGlyProAlaglySerlys	354
QY	1120	GGAGACATT-----AAGGGCATCCCTGGCATTTGTGTGHAAGCGTGGTCCCACATGCC	1067
DV			
Db	355	GlygluSerGlyAsnLysGlyGluProgly--SerAlaGlyAlaGlnGlyProPro---	372
QY	1066	CTTCCTCGGGTGATGTATGCACCAGGCTCGAACGTCGAGATGGGGCAAGAATTCATCCATCCCAC	1007
DV			
Db	373	-----GlyProserGlyGluGlnGlyLeuArgGlyProAsnGlyGlyGlyGlyGly	386
QY	1006	GAATTCATACATTCGGGAAGTCTCGGCACACAGGGGCTCGAGCCAGGTCCTCATAGC----	951
DV	:::	:::	:::
Db	386	luAlaGlySerAlaGlyProserGlyProProGlyLeuArgGlySerProglySerargg	406
QY	950	-----GCACAACAGGTACGTT---CCTGAGGGCGCTTGGGCAAGGA	911
DV			
Db	406	lyLeuProGlyAlaAspGlyProAlaGlyValMetGlyProProGlyProArgGlyAlat	426
QY	910	CTTGA---TGSTCTTGAGATCTCCAGCTGGCTTTGGCAGATGACCTGCATCACATAGTA	854
DV			
Db	426	hrGlyProAlaGlyValargGlyProAsnGlyAspserGlyArgProGlyGluProGlyL	446
QY	853	GGGTTGTCCTCCCTCTTTGAGTTTTGCTCATGTGCCCCATCAATCGCACTGTCAAT	794
DV			
Db	446	euMetGlyProArgGly-----PheProGlyAlaProGlyAsnValGlyProAlaG	463
QY	793	CATGAGATCTCCCTTTGTGCGTTCTCGGAACGGAACACGCCCGCGGGT-----	744
DV			
Db	463	lyLysGlnGlyProMetGlyLeuProGlyIleAspGlyArgProGlyProIleGlyProA	483
QY	743	-----CCGGACACAGGTCACGATATGCAG	719
DV			
Db	483	laGlyAlaArgGlyGluProGlyAsnIleGlyPheProGlyProlysglyProthrGlyA	503
QY	718	GTTGAGGGAGGGTGCTTTCAGCAGCGGGTAGAGGACTGCAGGTGAAGACGCCACCT-	660
DV			
Db	503	spProGlyLyseAsnGlyAspLysGlyHisAlaGlyLeuAlaGlyAlaalaArgGlyAla	523
QY	659	-----CCTTGAGCACACGTCGGCTGTAGGACGGCA	629
DV			
Db	523	lyProAspGlyAsnAsnGlyAlaGlnGlyProProGlyProGlnGlyValGlnGlyGlyL	543
QY	628	GG-----CCTTCTCCACCACTT-----CAAAGGGCTGTTGACT	596
DV	:::	:::	:::

Search completed: May 6, 2004, 11:02:17
Job time : 55.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:55:53 ; Search time 104 Seconds
(without alignments)
12396.225 Million cell updates/sec.

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattccattgttggtgta.....tgggattcctaaaaaaaaa 2043

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cpn2.1/USPTO_spool_p/US09645078/runat_06052004_104559_21739/app_query.fasta_1.2183
-DB=SPTRMBL_25 -QWTF=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09645078 @CNC 1 1 112 @runat_06052004_104559_21739 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2039	54.5	386	4 Q9Y5R3	Q9Y5R3 homo sapien

ALIGNMENTS		PRELIMINARY;		PRT;		386 AA.	
RESULT 1		Q9Y5R3		Q9Y5R3		Q9Y5R3	
ID		Q9Y5R3		Q9Y5R3		Q9Y5R3	
AC		Q9Y5R3		Q9Y5R3		Q9Y5R3	
DT		01-NOV-1999 (TRENBLrel. 12, Created)		01-NOV-1999 (TRENBLrel. 12, Created)		01-NOV-1999 (TRENBLrel. 12, Created)	
DT		01-NOV-1999 (TRENBLrel. 12, Last sequence update)		01-NOV-1999 (TRENBLrel. 12, Last sequence update)		01-NOV-1999 (TRENBLrel. 12, Last sequence update)	
DT		01-JUN-2003 (TRENBLrel. 24, Last annotation update)		01-JUN-2003 (TRENBLrel. 24, Last annotation update)		01-JUN-2003 (TRENBLrel. 24, Last annotation update)	
DE		N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand		N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand		N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand	
DE		sulfotransferase GST-3).		sulfotransferase GST-3).		sulfotransferase GST-3).	
GN		GST3.		GST3.		GST3.	
OS		Homo sapiens (Human)		Homo sapiens (Human)		Homo sapiens (Human)	
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX		NCBI_TaxID=9606;		NCBI_TaxID=9606;		NCBI_TaxID=9606;	
RN		[1]		[1]		[1]	
RP		SEQUENCE FROM N.A.		SEQUENCE FROM N.A.		SEQUENCE FROM N.A.	
RC		TISSUE=Tonsil;		TISSUE=Tonsil;		TISSUE=Tonsil;	
RX		MEDLINE=99264336; PubMed=10330415;		MEDLINE=99264336; PubMed=10330415;		MEDLINE=99264336; PubMed=10330415;	
RA		Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,		Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,		Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,	
RA		Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.,		Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.,		Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.,	
RT		"Sulfotransferases of two specificities function in the reconstitution		"Sulfotransferases of two specificities function in the reconstitution		"Sulfotransferases of two specificities function in the reconstitution	
RT		of high endothelial cell ligands for L-selectin."		of high endothelial cell ligands for L-selectin."		of high endothelial cell ligands for L-selectin."	
RL		J. Cell Biol. 145:899-910(1999).		J. Cell Biol. 145:899-910(1999).		J. Cell Biol. 145:899-910(1999).	
RN		[2]		[2]		[2]	
RP		SEQUENCE FROM N.A.		SEQUENCE FROM N.A.		SEQUENCE FROM N.A.	

Q8NCG5 homo sapien
Q8IV46 homo sapien
Q8LII1 mus musculu
Q8WUE5 mus musculu
Q8GZX3 homo sapien
Q8GUX4 mus sapien
Q8GZS9 homo sapien
Q8EP78 mus musculu
Q8NBO mus musculu
Q8W67 homo sapien
Q8N84 homo sapien
Q8N276 mus musculu
Q8OW3 mus musculu
Q8UED5 homo sapien
Q8V4C5 homo sapien
Q8EGC0 mus musculu
Q8316 homo sapien
Q8819 mus musculu
Q88199 mus sapien
Q8Z12 rattus norv
Q83403 torpedo cal
Q8VNC3 drosophila
Q8DOK5 mus musculu
Q8VNC4 drosophila
Q8MZ1 drosophila
Q851A8 cryptotagus
Q8V851 homo sapien
Q8V84 rhizobium m
Q8UZE1 cercopithec
Q8872 myxococcus
Q8SP4 bos taurus
Q81396 homo sapien
Q810J9 mus sapien
Q86045 homo sapien
Q8N473 homo sapien
Q8TTS5 bos taurus
Q8GFF2 streptomyce
Q85001 globodera p
Q8438 chironomus
Q8GK9 herpesvirus
Q8G65 mus musculu
Q8SP4 bos taurus
Q85000 rattus norv
Q86XJ2 homo sapien

RC	TISSUE=Tonsil;	
RA	Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,	
RA	Rosen S.D.,	
PT	"Chromosomal Localization and Genomic Organization for the	
RT	Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-	
RT	Sulfotransferase Gene Family.;"	
RL	Glycobiology 0:0-0(2001).	
RL	[3]	
RN	SEQUENCE FROM N.A.	
RP	MEDLINE=21332592; PubMed=11439191;	
RX	Yeh J.-C., Hiracka N., Petryniak B., Nakayama J., Ellies L.G.,	
RA	Rabuka D., Hindsaugh O., Marth J.D., Lowe J.B., Fukuda M.,	
RT	"Novel sulfated lymphocyte homing receptors and their control by a	
RT	core1 extension beta1,3-N-acetylglucosaminyltransferase.;"	
RL	Cell 105:957-969(2001).	
DR	EMBL; AF131235; AAD33015.1; -	
DR	EMBL; AF280088; AAG48246.1; -	
DR	EMBL; AF149783; AAK48417.1; -	
DR	GO; GO:0008146; P:sulfotransferase activity; TAS.	
DR	GO; GO:0007155; P:cell adhesion; TAS.	
DR	GO; GO:0006928; P:cell motility; TAS.	
DR	GO; GO:0007267; P:cell-cell signaling; TAS.	
DR	GO; GO:0006955; P:immune response; TAS.	
DR	GO; GO:0006477; P:protein amino acid sulfation; TAS.	
DR	InterPro; IPR001092; HLH_basic.	
DR	InterPro; IPR000863; Sulfotransferase.	
DR	Pfam; PF00685; Sulfotransfer; 1.	
DR	PROSITE; PS00038; HLH_1; 1.	
KW	Lectin; Selectin; Transferase.	
SQ	SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;	
Alignment Scores:		
Pred. No.:	1,53e-166	Length: 386
Score:	2038.00	Matches: 386
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	54.46%	Indels: 0
DB:	4	Gaps: 0
US-09-645-078-1 (1-2043) x Q9Y5R3 (1-386)		
Qy	128	ATGCTACTGCTCTAAATAAAGCTCTGCTGTTCTCTGTTCCAGATGGCATCTTG 187
Db	1	MetLeuLeuProLysLysMetLysLeuLeuLeuPheLeuValSerGlnMetAlaLeu 20
Qy	188	GCTCTATTCTCCACATAGCAGCACAAATCATGCTCCCTGTCTATGAAGGCACAGCC 247
Db	21	AlaLeuPhePheHisMetTyrSerHisAsnLysSerSerLeuSerMetLysAlaGlnPro 40
Qy	248	GAGCGCATGACGCTGCTGCTCTCTCTCTGCGCTCTGCTCTTTTGTGGGCGAG 307
Db	41	GluArgMetHisValLeuValLeuSerSerTyrArgSerGlySerSerPheValGlyGln 60
Qy	308	CTTTTGTGGCAGCACCAGATGTTTCTACCTGATGAGCCCGCTGCGAGTGTGGATG 367
Db	61	LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTyrHisValTyrMet 80
Qy	368	ACCTTCAGCAGAGCAGCCGCTGGATGTGCATGCTGCGGGATCTGATACGGGCC 427
Db	81	ThrPheLysGlnSerThrAlaTyrMetLeuHisMetAlaValArgAspLeuLeuArgAla 100
Qy	428	CTCTCTTGTGGCATGAGCTCTTTCATGCTACATGGAACCTGGTCCCGGAGACAG 487
Db	101	ValPheLeuLysCysAspMetSerValPheAspAlaTyrMetGluProGlyProArgArgGln 120
Qy	488	TCCAGCCTCTTTTACGTGGAGAACAGCCGGCCCTGTGTTCTGCACTCCGCTGTGACATC 547
Db	121	SerSerLeuPheGlnTyrGluAsnSerArgAlaLeuLysSerAlaProAlaCysAspLeu 140
Qy	548	ATCCACACAGATGAATCATCCCCGCTCTCATGCTGAGGCTCTGCTGAGTCACAGGCC 607
Db	141	IleProGlnAspGluLeuIleProArgAlaHisCysArgLeuLeuLysSerGlnGlnPro 160

Qy	608	TTTAGGTGTGGAGAGGCTGCGCTCTCTACAGCCAGCTGCTGTCTCAAGGAGGTGCGC 667
Db	161	PheGluValValGluLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
Qy	668	TTCTTCAACCTGACAGTCCCTCTACCCGCTGCTGAAAGACCCCTCCCTCAACCTGCATATC 727
Db	191	PhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsnLeuHisIle 200
Qy	728	GTGCACCTGCTCGGAGACCCCGGCGCTGTCGTTCCGAGACGCAACGCAAGGAGAT 787
Db	201	ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgTyrLysGlyAsp 220
Qy	788	CTCATGATTGACAGTCGCTGATGGGCGAGCATGAGCAAACTCAAGAGGAGGAC 847
Db	221	LeuMetLysAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAsp 240
Qy	848	CAACCTTACTATGTATGTCAGGTCTATGTCGAAAGCCAGCTGGAGATCTCAAGACCATC 907
Db	241	GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluLysLysThrIle 260
Qy	908	CAGTCTTGTCCCAAGGCGCTGACGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 967
Db	261	GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla 280
Qy	968	CGAGCCCTGTGGCCAGACTTCCGAAATGTATGAATTCGTTGGGATGGAAATTTTGGCC 1027
Db	281	ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro 300
Qy	1028	CATCTTCAGACTGCTGGTGCATACATACCCGAGGCAAGGCGATGGGTGACACGCTTTC 1087
Db	301	HisLeuGlnThrTyrPheValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
Qy	1088	CACACAAATGCCAGGATGCTCTTAATGTCTCCAGGCTGGCGCTGCTGCTGCTGCTGCT 1147
Db	321	HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTyrArgTyrSerLeuProTyr 340
Qy	1148	GAAAGGTTTCTCGACTTCAGAAAGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1207
Db	341	GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrArg 360
Qy	1208	CACGTCAGATCTGAACAAAGACAGAAACCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
Db	361	HisValArgSerGluGlnGluArgAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 380
Qy	1268	GTCCCTGAGCAATCCAC 1285
Db	381	ValProGluGlnIleHis 386
RESULT 2		
Q8NCG5	PRELIMINARY;	PRT; 386 AA.
ID	Q8NCG5	PRELIMINARY;
AC	Q8NCG5;	01-OCT-2002 (Tremblrel. 22, Created)
DT	01-OCT-2002 (Tremblrel. 22, Last sequence update)	
DT	01-JUN-2003 (Tremblrel. 24, Last annotation update)	
DE	Hypothetical protein FLJ90265.	
OS	Homo sapiens (Human)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	SEQUENCE FROM N.A.
RP	Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,	
RA	Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,	
RA	Yamanoto J., Makamatsu A., Nakamura Y., Kojima S., Nagahari K.,	
RA	Masuno Y., Oho T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,	
RA	Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;	
RT	"NEDO human cDNA sequencing project.;"	
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AK074746; BAC11177.1; -	
DR	GenBank; HGNC:1972; CHST4.	
DR	GO; GO:0008146; P:sulfotransferase activity; IEA.	
DR	InterPro; IPR000863; Sulfotransferase.	

DR	Pfam; PF00685; Sulfotransferase; 1.
KW	Hypothetical protein.
SQ	SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;
Alignment Scores:	
Pred. No.:	2,77e-166 Length: 386
Score:	2035.00 Matches: 385
Percent Similarity:	100.00% Conservative: 1
Best Local Similarity:	99.74% Mismatches: 0
Query Match:	54.38% Indels: 0
DB:	4 Gaps: 0
US-09-645-078-1 (1-2043) x Q8NCG5 (1-386)	
QY	128 ATGCTACTGCCTTAAATAAGACCTCGCTGTTCGGTTCCACATGCCATCTTG 187
Db	1 MetLeuLeuProLysMetLysLeuLeuPheLeuValSerGlnMetAlaLeu 20
QY	188 GCCTATTCTCCACATGTACAGCCACAACATCAGCTCCCTGCTCATGAAGCACAGCCC 247
Db	21 AlaLeuPheHisMetTyrrSerHisAsnIleSerSerLeuSerMetLysAlaGlnPro 40
QY	248 GAGCGATGCATGCTGCTGTCTCTCTCGCGCTCTGGCTCTCTCTTTTGTGGGGCAG 307
Db	41 GluArgMetHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGlyGln 60
QY	308 CTTTTTGGGACACACCAGATGTTTCTACCTGATGGAGCCCGCTGCAGCGTGTGGATG 367
Db	61 LeuPheGlyGlnHisProaspValPheTyrrLeuMetGluProAlaIrrPheHisValTrpMet 80
QY	368 ACCTCAAGCAGACCCCGCTGGATGCTGCATGGCTGTGCGGGATCTGATACGGGCC 427
Db	81 Thr-PheLysGlnSerThrAlaTrpMetLeuHisMetAlaValArgAspLeuIleArgAla 100
QY	428 GTCTTCTTGGCACATGAGCGCTTTTGTGCTAGCTACATGGACCTGTCCCGGAGCAG 487
Db	101 ValPheLeuCysaspMetSerValPheAspAlaTyrrMetGluProGlyProArgArgGln 120
QY	488 TCACAGCTCTTTTCAGTGGGAGAACACCGGGCCCTGTGTTCGCACCTCGCTGTGCATC 547
Db	121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuCysSerAlaProAlaCysaspIle 140
QY	548 ATCCACAAGATGAATATCCCCGGGTCTACTGAGCGCTCTGTGAGTCAACAGGCC 607
Db	141 IleProGlnAspGluIlelleProArgAlaHisCysArgLeuLeuCysSerGlnGlnPro 160
QY	608 TTTGAGTGTGTGAGAGGCGTCCGCTCTACAGCAGCTGTGTCTCAAGAGGTGGC 667
Db	161 PheGluValValGluLysAlaCysArgSerTyrrSerHisValLeuLysGluValArg 180
QY	668 TTCTTCACTGCAGTCCCTCTACCCGCTGCTGAAAAGACCCCTCCCTCAACTGCATATC 727
Db	181 PhePheAsnLeuGlnSerLeuTyrrProLeuLeuLysAspProSerLeuAsnLeuHisile 200
QY	728 GTGCACCTGTCCGGACCCCGGGCGGTGTTCCGTTCCCGAGAAGCAGCAAAGGGAGAT 787
Db	201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThryrLysGlyAsp 220
QY	788 CTCATGATTGCATCCGATTTGATGGGCGACATGAGCAAAAACTCAAGAGGAGGAC 847
Db	221 LeuMetIleAspAsnArgIleValMetGlyGlnHisGluGlnLysLeuLysGluAsp 240
QY	848 CAACCTCTATGTGATGAGGTTCATCTGCCAAGCCAGCTGGAGATCTCAAGACCATC 907
Db	241 GlnProTyrrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrrLysThrile 260
QY	908 CAGTCCCTTGGCCAAGGCCCTCGAGGAACGCTACTGCTGTGCGCTATGAGACCTGGCT 967
Db	261 GlnSerLeuProLysAlaLeuGlnGluArgTyrrLeuLeuValArgTyrrGluaspleula 280
QY	968 CGAGCCCTGTGGCCAGACTTCCCGAATGTATGAATTCGTGGGATGTGAATTCCTGCC 1027
Db	281 ArqAlaProValAlaGlnThrSerArgMetTyrrGluPheValGlyLeuGluPheLeuPro 300

Db	61	HisValTrpMetThrPheLysGlnSerThrAlaTrpMetLeuHisMetAlaValArgAsp	80
Qy	416	CTGATAGCGGCGCTTCTTGTGGACATGAGCGTCTTTCATCCCTACATGGAACCTGCT	475
Db	81	LeuileArgAlaValPheLeuLysCysAspMetSerValPheAspAlaTrpMetGluProGly	100
Qy	476	CCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAGAACACAGCCGGCCCTGTCTTCTGCACCT	535
Db	101	ProArgGlnSerSerLeuPheGlnTrpGluAsnSerArgAlaLeuLysSerAlaPro	120
Qy	536	CCCTGTGACATCCACAGATGAATCATCCCCGGCTCCTCAGCGCTCCCTGTC	595
Db	121	AlaCysAspIleIleProGlnAspGluIleIleProArgAlaHisCysArgLeuLeuLys	140
Qy	596	AGTCAACAGCCCTTGTAGCTGGTGGAGAGCCCTCCCTACAGCCACGCTGGTCTC	655
Db	141	SerGlnGlnProPheGluValValGluLysAlaCysArgSerTyrSerHisValValLeu	160
Qy	656	AAGAGGTGCGCTTCTTCAACTGCGAGTCCCTCTACCGCTGCTGAAAGACCCCTCCCTC	715
Db	161	LysGluValArgPhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeu	180
Qy	716	AACCTGTCATATCGTCACCTGTCGGGACCCCGGCGCTGTTCCGTCCCGAGAACGC	775
Db	181	AsnLeuHisIleValHisLeuValArgAspProArgAlaValPheArgSerArgGluArg	200
Qy	776	ACAAAGGAGATCTCATGATTCACAGTCGCGATTGTGATGGGCGAGCATGAGCAAACTC	835
Db	201	ThrLysGlyAspLeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeu	220
Qy	836	RAGAGGAGGACCAACCTTACTATGTATGATGAGGTCACTGCGCAAGCCACTCGAGATC	895
Db	221	LysLysGluAspGlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIle	240
Qy	896	TACAAGACCATCCAGTCTCTGCCAAGCCCTGCGAGAACCTACCTCTTGTGCGCTAT	955
Db	241	TyrLysThrIleGlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyr	260
Qy	956	GAGGACCTGCTGAGCCCTCTGCGCCAGACTTCCGAAATGATGAATTCGTGGATTG	1015
Db	261	GluAspLeuAlaArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeu	280
Qy	1016	GAATTCCTCCCATCTTCAGACCTGGGTGCATAACATCACCCGAGGCAAGGCGATGGT	1075
Db	281	GluPheLeuProHisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGly	300
Qy	1076	GACCAGCTTTCACACAATGCCAGGATGCCCTTAATGTCCTCCAGGCTGGCGCTGG	1135
Db	301	AspHisAlaPheHisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrp	320
Qy	1136	TCTTTCCTTATGAAAGGTTTCTCGACTTCAGAAAGCCCTGTGGCGATGCCATGAATTG	1195
Db	321	SerLeuProTyrGluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeu	340
Qy	1196	CTGGCTACCGCACCTCAGATCTGAACAGAACAGAGAACTGTTCTGCTGATCTTCTG	1255
Db	341	LeuGlyTyrArgHisValArgSerGluGlnGluArgAsnLeuLeuAspLeuLeu	360
Qy	1256	TCCTACCTGGACTGCTCCCTGAGCAAAATCCAC	1285
Db	361	SerThrTrpThrValProGluGlnIleHis	370
RESULT 4			
Q9R111 PRELIMINARY; PRT; 388 AA.			
AC Q9R111; 01-MAY-2000 (TrEMBLrel. 13, Created)			
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE L-selectin ligand sulfotransferase.			
GN CHS74.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99361934; PubMed=10435581;		
RA	Hiroaka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,		
RA	Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.,		
RT	"A novel, high endothelial venule-specific sulfotransferase expresses		
RT	6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34,"		
RL	Immunity 11:79-89(1999).		
DR	EMBL; AF109155; AAD45579.1; -.		
DR	MGD; MGI:1349479; Chst4.		
DR	GO; GO:0005194; F:cell adhesion molecule activity; IEA.		
DR	GO; GO:0005529; F:sugar binding; IEA.		
DR	GO; GO:0008146; F:sulfotransferase activity; IEA.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0007157; P:heterophilic cell adhesion; IEA.		
DR	InterPro; IPR000863; Sulfotransferase.		
DR	Pfam; PF00685; Sulfotransfer; 1.		
KW	Lectin; Selectin; Transferase.		
SQ	SEQUENCE 388 AA; 44636 MW; 6D5371APB6884AEE CRC64;		
Alignment Scores:			
Pred. No.:	4.54e-122 Length: 388		
Score:	1521.50 Matches: 283		
Percent Similarity:	85.13% Conservative: 49		
Best Local Similarity:	72.56% Mismatches: 55		
Query Match:	40.66% Indels: 3		
DB:	11 Gaps: 2		
US-09-645-078-1 (1-2043) x Q9R111 (1-388)			
Qy	128	ATGCTACTGCCTAAATAAATGAAGCTCTCTGCTGTTTCTGTTTCCAGATGCCATCTTG	187
Db	1	MetMetLeuLysLysGlyArgLeuLeuMetPheLeuGlySerGlnValIleValVal	20
Qy	188	GCTCTATTCTTCACATGTATGACCCACACATCAGCTCCCTCTCTATGAGCACAGCC	247
Db	21	AlaLeuPheIleHisMetSerValHisArg-----HisLeuSerGlnArgGluLeuSer	38
Qy	248	GACGCG---ATGCACCTGCTGCTGCTCTCTCTGCGCTCTGCTCTCTTTTGTGGG	304
Db	39	ArgArgProValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGly	58
Qy	305	CAGCTTTTGGGAGACACCCAGATGTTTCTACCTGATGAGCCGCTGGCAGCTGTGG	364
Db	59	GlnLeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrp	78
Qy	365	ATGACCTTCAAGCAGACGACCGCTCGATGCTGCACATGCTGCGGATCTGATACGG	424
Db	79	MetThrPheThrSerThrAlaTrpLysLeuHisMetAlaValArgAspLeuLeuArg	98
Qy	425	GCCCTCTTCTTGTCCGACATGAGCGTCTTTGATGCTTACATGGAACCTGTCCTCCG	484
Db	99	SerValPheLeuLysCysAspMetSerValPheAspAlaTyrMetAsnProGlyProArgLys	118
Qy	485	CAGTCCAGCTCTTTCAGTGGGAGAACAGCCGCGCTGTTCTGCTCTGCTGCTGTGAC	544
Db	119	GlnSerSerLeuPheGlnTrpGluGlnSerArgAlaLeuLysSerAlaProValCysAsp	138
Qy	545	ATCATCCACACAGATGAATCATCCCCGGGCTCCTGTCAGGCTCTGTCAGTCAACAG	604
Db	139	PhePheProAlaHisGluIleSerSerProLysHisCysLysLeuLeuLysGlnGln	158
Qy	605	CCCTTTGAGTGGTGGAGAGGCTGCTGCTCTACAGCCACGCTGCTCAAGGAGGTG	664
Db	159	ProPheAspMetValGluLysAlaCysArgSerHisGlyPheValValLeuLysGluVal	178
Qy	665	CGCTTCTTCAACCTGAGTCCCTCTACCCGCTGCTGTAAGACCCCTCCCTCAACCTGCAT	724
Db	179	ArgPheLeuSerLeuGlnAlaLeuTyrProLeuLeuThrAspProSerLeuAsnLeuHis	198
Qy	725	ATCGTGACCTGCTGCTCCGGGACCCCGGCGCGTTCCTGTTCCCGAGAACCCACAAAGGA	784

```
Db 199 ValValHisLeuValA-gaspProArgAlaValPheArgSerArgGluHisThrThrile 218
Qy 785 GATCTCATGATGACAGTCGCGATGTCATGGGCGAGCATGAGCAAAACCTCAAGAAAGGAG 844
Db 219 GluLeuValValAspSerHisIleValLeuGlyGlnHisLeuGluThrIleLysGluGlu 238
Qy 845 GACCAACCTACTATGTCAGTCAGTCATGTCGCAAGCAGCTGGAGATCTACAGACC 904
Db 239 AspGlnProTyrTyrAlaMetLysIleLeuCysLysSerGlnValAspIleValLysAla 258
Qy 905 ATCCAGTCCTTGGCCCAAGCCCTGCAGGAACGCTACCTGCTGTGCGCTATGAGGACCTG 964
Db 259 IlegInThrLeuProGluAlaLeuGlnGlnArgTyrLeuPheLeuA-gTyrGluAspLeu 278
Qy 965 GCTGAGCCCTGTGGCCAGCACTCCCGAATGATGAATTCGTGGGATTCGAATCTTGG 1024
Db 279 ValArgAlaProLeuAlaGlnThrThrArgLeuTyrLysPheValGlyLeuAspPheLeu 298
Qy 1025 CCCATCTTCAGACTGGTGCATACATACCCAGGCGAGGCGATGGGTGACACAGCT 1084
Db 299 ProHisLeuGlnThrTyrValHisAsnValThrArgGlyLysGlyMetGlyGlnHisAla 318
Qy 1085 TTCCACAAATGCCAGGATGCCCTTAATCTCTCCAGGCTTGGCGCTGCTTTTCCCC 1144
Db 319 PheHisThrAsnAlaArgAsnAlaLeuAsnValSerGlnAlaTyrArgTyrSerLeuPro 338
Qy 1145 TATGAAGAGTTTCTCGACTTCAGAAACCTGTGGCGATGCGATGAATTCCTGGGCTAC 1204
Db 339 TyrGluLysValSerGlnLeuGlnAspAlaCysGlyGlnAlaMetAspLeuLeuGlyTyr 358
Qy 1205 CGCCAGCTCAGATCTGAACAGAGACAGAGAACCTGTGCTGGATCTTCTCTACCTGG 1264
Db 359 LeuGlnValArgSerGlnGlnGlnGlyAsnLeuSerLeuAspLeuLeuSerSer 378
Qy 1265 ACTGTCCTCAGCAATCCATAGAGGGT 1294
Db 379 HisIleLeuGlyGlnValPheArgGluGly 388

RESULT 5
Q9WUE5
ID Q9WUE5 PRELIMINARY; PRT; 388 AA.
AC Q9WUE5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin
DE 6/keratan) sulfotransferase 4).
GN CHS74.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,
RA Huang C.-C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,
RA Fleischmann W., Gaasterland T., Glassi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
```

```
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RL EMBL; AF131236; AAD33016.1; -.
DR EMBL; AK009113; BAB26078.1; -.
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Alignment Scores:
Pred. No.: 6,75e-122 Length: 388
Score: 1519.50 Matches: 283
Percent Similarity: 85.13% Conservative: 49
Best Local Similarity: 72.56% Mismatches: 55
Query Match: 40.61% Indels: 3
DB: 11 Gaps: 2

US-09-645-078-1 (1-2043) x Q9WUE5 (1-388)
Qy 128 ATGCTACTCCCTAAATAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGCCATCTG 187
Db 1 MetMetLeuLeuLysLysGlyArgLeuLeuMetPheLeuGlySerGlnValVal 20
Qy 188 GCTCTATTCTTCCATGATGACGACCAACATCAGCTCCCTGCTCTATGAGGCACAGCC 247
Db 21 AlaLeuPheIleHisMetSerValHisArg-----HisLeuSerGlnArgGluIleSer 38
Qy 248 GAGCGC---ATGCACGCTGCTGCTTCTGCTTCTTCTGCGCTCTGCTTCTTCTTCTGGG 304
Db 39 ArgArgProValHisValLeuValLeuSerSerTyrArgSerGlySerPheValGly 58
Qy 305 CAGCTTTTGGGAGCAGCCAGCATGTTTCTACCTGATGAGCCGCCCTGGCAGCTGGG 364
Db 59 GlnLeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTyrHisValTyr 78
Qy 365 ATGACCTTCAAGCAGACGACCGCTGATGCTGCATGCTGCGGATCTGATACCG 424
Db 79 MetThrPheThrSerSerThrAlaTyrLysLeuHisMetAlaValArgAspLeuLeuArg 98
Qy 425 GCGCTTCTTGTGCGACATGAGCGCTTTGTGATGCTTACATGGAACCTGCTGCCGGAGA 484
Db 99 SerValPheLeuCysAspMetSerValPheAspAlaTyrMetAsnProGlyProArgLys 118
Qy 485 CAGTCCAGCTCTTTCAGTGGGAGACACCGCGCTGCTGTTCTGCTACCTGCTGCTGAC 544
Db 119 GlnSerSerLeuPheGlnTyrGluGlnSerArgAlaLeuCysSerAlaProValCysAsp 138
Qy 545 ATCATCCCAACAGATGAATCATCCCGCGCTCTACTGACAGCTCTCTGCTGAGTCAACAG 604
Db 139 PhePheProAlaHisGluIleSerSerProLysHisCysLysLeuLeuCysGlyGlnGln 158
Qy 605 CCCTTTGAGGTGGTGAAGAAGGCTCCGCTCTCTACAGCCAGTGGTGGTCTCAGGAGGTG 664
Db 159 ProPheAspMetValGluLysAlaCysArgSerHisGlyPheValValLeuLysGluVal 178
Qy 665 CGCTTCTTCAACTGCGAGTCCCTCTACCGCTCTCTGAAAGACCCCTCCCTCCCTGAT 724
Db 179 ArgPheLeuSerLeuGlnAlaLeuTyrProLeuLeuThrAspProSerLeuLeuHis 198
```



```
Db 256 AlathrLeuLysPro---ProPheLeuArgGlyArgTyArgLeuValArgPheGlu 274
QY 959 GACCTGGCTCAGCCCTGTCGCCAGACTTCCCGAATGATGAATTTCGTGGGATTGGAA 1018
Db 275 AspLeuAlaArgGluProLeuAlaGluLeuArgAlaLeuTyAlaPheThrGlyLeuSer 294
QY 1019 TTCTGGCCCATCTTCAGACCTGGGTGCATTAACATCACCCGAGCAAGGGCATGGGT--- 1075
Db 295 LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 --GACCACGCTTCCACACAAATCCAGGAGTCCCTTAATGTCCTCCAGGCTTGGCGC 1132
Db 315 ArgArgGluAlaPheLeuThrSerArgAsnAlaLeuValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTGGCCCTAAGAAAGCTTTCGACTTCAGAAAGCTGTGGCGATGCCATGAAT 1192
Db 335 HisAlaLeuProPheAlaLysIleArgArgValGlnGluLeuCysAlaGlyAlaLeuGln 354
QY 1193 TTGTGGGCTACCCGACGTCAGATCTGAACCAAGACAGAGAAACCTGTCGTGATCTT 1252
Db 355 LeuLeuGlyTyArgProValTySerGluAspGluGlnArgAsnLeuAlaLeuAspLeu 374
QY 1253 CTGTCTACCTGGACTGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTTTGCTGCC 1312
Db 375 Val-----LeuPro-----ArgGlyLeuAsnGlyPhe----- 383
QY 1313 ACCTGGTGTGACCTCAGTCACTTCTCTGAATGCTTCTGAGCCT 1357
Db 384 ThrTrpAla-----SerSerThrAlaSerHisPro 393

RESULT 7
OSQUP4
ID Q9QUP4 PRELIMINARY; PRT; 395 AA.
AC Q9QUP4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN CHST5 OR i-GLCNAC-6-ST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Intestine;
RX MEDLINE=99423499; PubMed=10491328;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
RL Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
DR EMBL, AF176841; AAD56003.1; -
DR EMBL, AF176840; AAD56002.1; -
DR MGD, MGI:1931825; Chst5.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 395 AA; 44537 MW; - 3PFD71E43ED383BE CRC64;

Alignment Scores:
Pred. No.: 6,92e-79 Length: 395
Score: 1020.00 Matches: 221
Percent Similarity: 66.58% Conservative: 44
Best Local Similarity: 55.53% Mismatches: 120
Query Match: 27,26% Indels: 13
DB: 11 Gaps: 8

US-09-645-078-1 (1-2043) x Q9QUP4 (1-395)
QY 128 ATGCTACTGCCTAAA-----AAAATGAAGCTCCTGCTGTTTCTGTTTCCAGATGCC 181
```

```
Db 1 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuLeuMetValGlnThrGly 20
QY 182 ATCTTGGCTCTATTCTTCCACATGTACAGCCACACATCAGCTCCCTCTCTATGAAGCA 241
Db 21 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 36
QY 242 CAGCCCGAGCGCATCACTGCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 301
Db 37 LeuGlyGluArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheVal 56
QY 302 GGGCAGCTTTTGGGCGAGCACCAGATGTTTCTACCTGATGGAGCCCGCTGGCACGTG 361
Db 57 GlyGlnLeuPheSerGlnHisProAspValPheTyLeuMetGluProAlaTrpHisVal 76
QY 362 TGGATGACCTTCAAGCAGACACCCCTGGATGCTGTCACATGGCTGCGGGATCTGATA 421
Db 77 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeu 96
QY 422 CGGGCGCTTCTTGTGCGACATGAGCGTCTTTGATGCTACATCGAAGACCTGGTCCCGG 481
Db 97 ArgSerValPheLeuCysAspMetAspValPheAspAlaTyLeu---ProTrpArgArg 115
QY 482 AGACATGTCAGCTCTTTCAGTGGGAGAACAGCCGGGCTGTGTCTGACCTGCCTGTGT 541
Db 116 AsnIleSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProValCys 135
QY 542 GACATCATCCCAAGATGAATCATCCCGGGCTCACTGCAGGCTCTGTGCAAGTCAA 601
Db 136 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLysProLeuCysAlaThr 155
QY 602 CAGCCCTTTGAGTGGTGGAGAAAGCCCTGCTCCTACAGCCACGTGGTCTCAAGAG 661
Db 156 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTySerHisValValLeuLysGlu 175
QY 662 GTGCGCTTCTTCAACCTGCACTCCCTTACCCGCTGCTGAAAGACCCCTCCCTCAACCTG 721
Db 176 ValArgPheAsnLeuGlnValLeuTyProLeuLeuSerAspProAlaLeuAsnLeu 195
QY 722 CATATGTCACCTGTGTCGGGACCCCGGGCGCTGTTTCCGTTCCCGAAGACGCAACAAG 781
Db 196 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAla 215
QY 782 GAGATCTCATGATTCAGTGCCTGCTGATGGGCGAGCATGACCAAACTCAAGAAG 841
Db 216 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---Glu 234
QY 842 GAGGACCAACCTTACTATGTCGTGATGAGGTCTATGCCAAAGCCAGCTGGAGATCTACAAG 901
Db 235 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 254
QY 902 ---ACCATCAAGTCTTGGCCCAAGGCCCTGCGAGGAACGCTACTGCTGTGGCTATGAG 958
Db 255 AlaAlaLeuHisLysProProPheLeuGlnAspArgTyArgLeuValArgTyGlu 274
QY 959 GACCTGCTCGAGCCCTGTCGCCAGACTTCCCGAATGTATGAATTCGTGGATTGGAA 1018
Db 275 AspLeuAlaArgAspProLeuThrValIleArgGluLeuTyAlaPheThrGlyLeuGly 294
QY 1019 TTCTTGGCCCATCTTCAGACCTGGGTGCATAACATCACCCGAGCAAGGGCATGGGT--- 1075
Db 295 LeuThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCACGCTTCCACACAAATGCCAGGATGCCCTTAATGCTTCCAGGCTTGGCGC 1132
Db 315 ArgArgGluAlaPheLysThrThrSerArgAspAlaLeuSerValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTGGCCCTAAGAAAGCTTTCGACTTCAGAAAGCTGTGGCGATGCCATGAAT 1192
Db 335 HisThrLeuProPheAlaLysIleArgArgValGlnGluLeuCysGlyGlyAlaLeuGln 354
QY 1193 TTGCTGGGCTACCCGACGTCAGATCTGAACCAAGACAGAGAAACCTGTTGCTGATCTT 1252
```



```

QY 913 CTGCCCCAGCCCTGCAGGAACGCTACCTGCTGTGCGCTATGAGGACCTGGCTCGAGC 972
Db 359 yAlaProAlaTripleuArgArgTyrLeuArgLeuArgTyrGluAspLeuValTrpG1 379
QY 973 CCTGTGGCCAGACATCCGATGTATGAATTCGTGGGATTTGGATTTCTGCCCATCT 1032
Db 379 nProGlnAlaGlnLeuArgLeuArgLeuArgPheSerGlyLeuA-gThrLeuAlaAlaLe 399
QY 1033 TCAGACCTGGTGCATAACATACCCGAGGACGAGGCATGGGT--GACCAGCTTTCCA 1089
Db 399 uAspAlaPheAlaPheAsnMetThrArgGlySerAlaTyrGlyAlaAspArgProPheHi 419
QY 1090 CACAAATGCCAGGATGCCCTTAATGCTCCCGAGCTGGCGCTGCTTTCGCCATGA 1149
Db 419 sLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrpArgGluArgLeuSerGlnG1 439
QY 1150 AAAGTTTCCTCGATCTCAGAAACCTGTGGCGATGCCATGAATTTGCTGGGTACCGCCA 1209
Db 439 uGlnValArgGlnValGluThrAlaCysAlaProAlaMetArgLeuLeuAlaTyrProAr 459
QY 1210 CGTCAGATCTGAACAAGACAGAGAAACCTGTTG-----CTGGATCTTCT 1254
Db 459 gSerGlyAspGluArgAspArgLysThrValArgGluGlyGluThrProLeuGluThrLy 479
QY 1255 GTCTACCTGGACTGTC 1270
Db 479 sAlaAsnTrpAlaVal 484

RESULT 10
Q99NB0 PRELIMINARY; PRT; 484 AA.
AC Q99NB0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chondroitin 6-sulfotransferase-2.
GN CHST7 OR MC6ST-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
FX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Kitagawa H., Uyama T., Sugahara K.;
RT "Cloning and Expression of Mouse Chondroitin 6-sulfotransferase-2.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046929; BAB40372.1; -.
DR PIR; JC7350; JC7350.
DR MGD; MGI:1891767; Chst7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 484 AA; 54751 MW; 402C5E1ED185FDF8 CRC64;

Alignment Scores:
Pred. NO.: 3,09e-45 Length: 484
Score: 629.50 Matches: 154
Percent Similarity: 48.50% Conservative: 72
Best Local Similarity: 33.05% Mismatches: 176
Query Match: 16.82% Indels: 64
DB: 11 Gaps: 11

US-09-645-078-1 (1-2043) x Q99NB0 (1-484)
QY 19 TACGGAAGACGACAGAGGGTAGAGGAGAAAGCGATGGCCCGG---CTAGCAGTG--- 72
Db 35 HisSerGluGlnAspLysGlyArgAsnCysProGlyLeuGlnArgSerLeuGlyValTrp 54
QY 73 AGCTCTCAAAAGCAGCAGGAGGAGCCCAAGCCACAGGCTCTCCACTTCAGCACATGCT 132

```

```

Db 55 SerLeuGluAlaAlaAlaAlaGlyLeuArgGluGlnGlyAlaGluValArgSerLeuAla 74
QY 133 ACTGCCCTAAATAAATGAGAGCTCCTGCTTTCTGCTTCCAGATGCCATCTTGGCTCT 192
Db 75 -----GluGlyAsnProAspArgSerProGlySer 84
QY 193 ATTCTTCCACATGTACACCCACAAACATCAGCTCCCTGTCTATGAAAGGCACAGCCCGAGCG 252
Db 85 -----Pro-GlyAsnLeuSerAlaValGlyGluAlaValThrGlnGluLy 99
QY 253 CATGCAGCTGCTGCTGCTCTCTCTGCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 312
Db 99 sGlnHisIleTyrValHisAlaThrTrpArgThrGlySerSerPheLeuGlyGluLeuPhe 119
QY 313 TGGGAGCAGCCAGCATGTTTCTACCTGATGAGAGCCCGCTGGCACCTGTGTGGATGACCTT 372
Db 119 eAsnGlnHisProAspValPheTyrLeuTyrAspProMetTrpHisLeuTrpGlnAlaLe 139
QY 373 CAAGCAGACACCGCTCGATGCTGCACATGCTGCGGATGCTGATATGAGGCGGCTCT 432
Db 139 uTyrProGlyAspAlaGluSerLeuGlnGlyAlaLeuArgAspMetLeuArgSerLeuPhe 159
QY 433 CTTGTGCGACATGACGCTTGTGATGCCCTACATGAACCTGGTCCCGGAGACAG----- 487
Db 159 eArgCysAspPheSerValLeuArgLeuTyrAlaGlnProGlyAspProGlyGluArgAl 179
QY 486 -----TCCAGCGCTCTTTTCAGTGGAGAACACAGCCCGGCCCT 522
Db 179 aProAspSerAlaAsnLeuThrAlaMetLeuPheArgTrpArgThrAsnLysValI1 199
QY 523 GTGTTCTGCACCTGCTGTGACATCATCCCAAGATGA-----ATCATCCC 570
Db 199 eCysSerProProLeuCysProAlaAlaProAlaArgAlaAspValGlyLeuValG1 219
QY 571 CCGGGTCACTGCAGGCTCTGTGAGTCAACAGCCCTTTGAGGTGTGGAGAGGCGCTG 630
Db 219 uAspLysAlaCysGluSerThrCysProProValSerLeuArgAlaLeuGluAlaGluCy 239
QY 631 CGCTCTTACACCCAGCTGTGCTCAAGAGGTGGCTTCTTCACTGAGTCCCTCTA 690
Db 239 sArgLysTyrProValValIleLysAspValArgLeuAspLeuGlyValLeuVa 259
QY 691 CCGCTGTCTGAAAGACCCCTCCCTCAACCTCATCTGTGCACCTGTGTGGGAGCCCGG 750
Db 259 lProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheArgAspProAr 279
QY 751 GCGCTGTCTCTCCCTCCAGAACCGCAAGGGAGATCTCATGATTACAGT----- 802
Db 279 gAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgGluSerIleGlnVa 299
QY 803 -----CGATTGTGATG----- 814
Db 299 lLeuArgThrArgGlnArgGlyAspHisPheHisArgValLeuLeuAlaHisGlyValAs 319
QY 815 -----GGGACGATGAGCAAAACTCAAGAGGAGGACCACCTACTAT----- 859
Db 319 pAlaArgProGlyGlyGlnAlaArgAlaLeuProSerAlaProArgAlaAspPhePheLe 339
QY 860 -----GTGATGCAGGTCTATCTCCAAAGCCAGCTGGAGATCTACAGACCATCCAGTC 912
Db 339 uThrSerAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuLeuPheThrArgG1 359
QY 913 CTGCCCCAGGCGCTGCAGGAACGCTACTCTGTGCGCTATGAGGACCTGGCTCGAGC 972
Db 359 yAlaProAlaTrpLeuArgArgTyrLeuArgLeuArgTyrGluAspLeuValTrpG1 379
QY 973 CCTGTGGCCAGACCTCCCGAATGTATGAATTCGTGGGATTTGGATTTCTGCCCATCT 1032
Db 379 nProGlnAlaGlnLeuArgArgLeuArgPheSerGlyLeuArgThrLeuAlaAlaLe 399
QY 1033 TCAGACCTGGTGCATAACATCACCAGGAGGAGGCGATGGGT---GACCAGCTTTCCA 1089
Db 399 uAspAlaPheAlaPheAsnMetThrArgGlySerAlaTyrGlyAlaAspArgProPheHi 419

```

```

QY 1090 CACAAATCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGGTCTTGGCCCTATGA 1149
Db 419 sLeuSerAlaArgAspAlaAaGluAlaValHisAlaTrpArgGluArgLeuSerGlnG1 439
QY 1150 AAAGCTTCTCGACTTCAGAAAGCTGGCGATGCCATGAATTGCTGGCTACCCGCA 1209
Db 439 uGlnValArgGlnValGluThrAlaCysAlaProAlaMetArgLeuLeuAlaTrpProAr 459
QY 1210 CGTCAGATCTCAACAAGAACAGAGAAACCTCTTG-----CTGGATCTTCT 1254
Db 459 gSerGlyAspGluArgAspArgGlyThrValArgGluGluThrProLeuGluThrLy 479
QY 1255 GTCTACCTGGAGTCTC 1270
Db 479 sAlaAsnTrpAlaVal 484

RESULT 11
O75667
ID O75667 PRELIMINARY; PRT; 486 AA.
AC O75667
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (Carbohydrate
DE (N-acetylglucosamine 6-O) sulfotransferase 7).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Uchimura K., Muramatsu T.;
RT Identification and molecular characterization of a cDNA encoding a
RT novel N-acetylglucosamine-6-O-sulfotransferase."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB040711; BAB13770.1; -;
DR EMBL; BC045537; AAH45537.1; -;
DR PIR; JC7351; JC7351.
DR Genew; HGNC:13817; CHST7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 486 AA; 54266 MW; 3F1FD1430B3C8E95 CRC64;

Alignment Scores:
Pred. No.: 4, 97e-44 Length: 486
Score: 615.50 Matches: 155
Percent Similarity: 48.22% Conservative: 62
Best Local Similarity: 34.44% Mismatches: 163
Query Match: 16.45% Indels: 70
DB: 4 Gaps: 12

US-09-645-078-1 (1-2043) x O75667 (1-486)

QY 7 CATTCTGTGGTACGGAGACGACAGAGGGTAGAGGAAAGCGCATGGCCGGCTA 66
Db 44 HisCysProGlyLeuGlnArg-----SerLeu 52
QY 67 GCAGTG---AGCCTCTCAAAAGCAGCAGGAGCCCAAGCCCAAGGCTTCCACTTCAG 123
Db 53 GlyValTrpSerLeuGluAlaAlaAlaGlyGluArgGluGluGluGluAlaArg 72
QY 124 CACAATGCTACTGCTCTAAATAAAGCTCTCTGTTCTTGGTTTCCAGATGGCCAT 183

```

```

Db 73 AlaAlaGluGluGlyGlyAlaAsnGlnSerProArg-----PheProSerAsnLeu 89
QY 184 CTTGGCTCTATTCTTCCACATGTACAGCCACACATCAGCTCCCTGTCTATGAAGGCACA 243
Db 90 SerGlyAlaValGlyGluAlaVal-SerArg----- 99
QY 244 GCCCGAGCCATGCAGCTGCTGGTCTCTCTCTGGCGCTCTGGCTCTCTCTTTTGTGG 303
Db 100 ----GluLysGlnHisIleTyValHisAlaThrTrpArgThrGlySerSerPheLeuG1 118
QY 304 GCAGCTTTTGGCGACGACCCAGATGTTTCTACCTGTAGTGAGCCCGCTCGCAGCTGTG 363
Db 118 yGluLeuPheAsnGlnHisProAspValPheTyLeuTyGluProMetTrpHisLeuTr 138
QY 364 GATGACCTTCAAGCAGAGACCCGCTGGATGCTGCATGCTGGTGTGGGATCTCATAG 423
Db 138 pGlnAlaLeuTyProGlyAspAlaLeuSerLeuGlnGlyAlaLeuArgAspMetLeuAr 158
QY 424 GGCGCTCTTCTTGTGGACATGAGGCTCTTTGATGCCTACATGGAACCTGGT---CCCG 480
Db 158 gSerLeuPheArgCysAspPheSerValLeuArgLeuTyAlaProProGlyAspProAl 178
QY 481 GAGACAG-----TCCAGCCTCTTTTCAGTGGGGAACAG 513
Db 178 aAlaArgAlaProAspThrAlaAsnLeuThrThrAlaAlaLeuPheArgTrpArgThrAs 198
QY 514 CGGGCCCTGTGTCTGCACCTGCTGTGACATCATCCACAGATGAA----- 562
Db 198 nLysValIleCysSerProProLeuCysProGlyAlaProArgAlaArgAlaGluValG1 218
QY 563 -ATCATCCCCGGGCTCACTGCAGGCTCTGTGCAGTCAACAGCCCTTTGAGTGGTGA 621
Db 218 yLeuValGluAspThrAlaCysGluArgSerCysProValAlaIleArgAlaLeuG1 238
QY 622 GAAGGCTCCCTCTCTACAGCAGCTGTGTCTCAAGAGGTGCGCTTCTTCAACTGCA 681
Db 238 uAlaGluCysArgLysTyProValValIleLysAspValArgLeuLeuAspLeuG1 258
QY 682 GTCCTCTACCCGCTCTGAAAGACCCCTCCCTCAACCTGCATATCGTCACCTGGTCCG 741
Db 258 yValLeuValProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheAr 278
QY 742 GAGCCCGCGGCGCTTCCGTTCCGAGAACGACGACAAAGGAGATCTCATGATTGACAG 801
Db 278 gAspProArgAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgLysLe 298
QY 802 T-----CGCATTTGTGATG----- 814
Db 298 rIleGlnValLeuArgThrArgGlnArgGlyAspArgPheHisArgValLeuLeuAlaHi 318
QY 815 -----GGCGAGCATGAGCAAAAACCTCAAGAGGAGGACCAACCCTA 855
Db 318 sGlyValGlyAlaArgProGlyGlyGlnSerArgAlaLeuProAlaAlaProArgAlaAs 338
QY 856 CTAT-----GTGATGCGGTCTATCTGCCAAGCCAGCTGGAGATCTACAAGAC 903
Db 338 pPhePheLeuThrGlyAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuLeuPh 358
QY 904 CATCCAGTCTTGGCCCAAGCCCTGCAGGAACGCTACTGCTGTGTGGCTATGAGGACT 963
Db 358 eAlaArgGlyAlaProAlaTrpLeuArgArgTyLeuArgLeuArgTyGlyLysAspLe 378
QY 964 GGCTCCAGCCCTGTGGCCCGACACTTCCCGATGATGAATTCGTGGGATTCGAATTC 1023
Db 378 uValArgGlnProArgAlaGlnLeuArgLeuLeuArgPheSerGlyLeuArgAlaLe 398
QY 1024 GCCCCATCTTCAGACTGGGTGCATCAATCAATCAACCCGAGCAAGGCGCATGGGT---GACCA 1080
Db 398 uAlaAlaLeuAspAlaPheAlaLeuAsnMetThrArgGlyAlaAlaTyGlyAlaAspAr 418
QY 1081 CGCTTTCCACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTTGGCGCTGGTCTTT 1140
Db 418 sProPheHisLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrpArgGluArgLe 438

```

QY 1141 GCCTATGAAAAGGTTCTCGACTTCAGAAAGCCTGTGGCGATGCCATGAATTTGCTGGG 1200
Db 438 userArgGluGlnValArgGlnValGluAlaCysAlaProAlaMetArgLeuLeuAl 458
QY 1201 CTACGCCACGTCAGATCTGAACAGAA 1228
Db 458 aTyr-----ProArgSerGlyGluGlu 465

RESULT 12
Q9NS84 PRELIMINARY; PRT; 486 AA.
AC Q9NS84;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Chondroitin 6-sulfotransferase-2.
GN CST-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20347214; PubMed=10781596;
RA Kitagawa H., Fujita M., Ito N., Sugahara K.;
RT "Molecular cloning and expression of a novel chondroitin 6-O-
sulfotransferase.";
RL J. Biol. Chem. 275:21075-21080(2000).
DR EMBL: AB037187; BAB03217.1; -.
DR GO: GO:0016021; C: integral to membrane; TAS.
DR GO: GO:0008459; F: chondroitin 6-sulfotransferase activity; TAS.
DR GO: GO:0005976; P: polysaccharide metabolism; TAS.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransfer; 1.
KW Transferrase.
SQ SEQUENCE 486 AA; 54294 MW; B524D1430B3C9075 CRC64;

Alignment Scores:
Pred. No.: 6.06e-44 Length: 486
Score: 614.50 Matches: 155
Percent Similarity: 48.22% Conservative: 62
Best Local Similarity: 34.44% Mismatches: 163
Query Match: 16.42% Indels: 70
DB: 4 Gaps: 12

US-09-645-078-1 (1-2043) x Q9NS84 (1-486)
QY 7 CATTGTGTGGTACGGAGACGACAGAAAGGTTAGAGAGAAAGCGCGTGGCGGCTA 66
Db 44 HisCysProGlyLeuGlnArg-----SerLeu 52
QY 67 GCAGTG---AGCCTCAAAAGCAGCGAGGAGCCCAAGCCACAAAGGTCTTCCACTTCAG 123
Db 53 GlyValTrpSerLeuGluAlaAlaAlaGlyGluArgGluGlnGlyAlaGluAlaArg 72
QY 124 CACAATGCTACTGCCTAAATAAATGAAGTCTCTGCTGTTTCCAGATGGCCAT 183
Db 73 AlaAlaGluGluGlyGlyAlaAsnGlnSerProArg-----PheProSerAsnLeu 89
QY 184 CTGGCTTATCTTCCAGTGACGACGACCAACATCAGCTCCCTGCTATGAGGACACA 243
Db 90 SerGlyAlaValGlyGluAlaVal-SerArg----- 99
QY 244 GCCCGAGCGCATGACGCTGCTGTTCTCTCTCTCTGGGCTCTGGCTCTTTTGGGG 303
Db 100 ----GluGlyGlnHisIleTyrValHisAlaThrTrpArgThrGlySerSerPheLeuGlu 118
QY 304 GCAGCTTTTGGGACGACCCAGATGTTTCTACCTGATGGAGCCCGCTGGCAGCTGTG 363
Db 118 yGluLeuPheAsnGlnHisProAspValPheTyrLeuTyrGluProMetTrpHisLeuTr 138

QY 364 GATGACCTTCAAGCAGACCGCCTGGATGTCACATGGCTGCGGATCTGATACG 423
Db 138 pGlnAlaLeuTyrProGlyAspAlaGluSerLeuGlnGlyAlaLeuArgAspMetLeuAr 158
QY 424 GCGCGCTTCTTGTGGACATGAGCCTCTTGTATGCTTACATGACGAACTGGT---CCCGC 480
Db 158 sSerLeuPheArgCysAspPheSerValLeuArgLeuTyrAlaProProGlyAspProAl 178
QY 481 GAGACAG-----TCCAGCCTCTTTCAGTGGGAGAACAG 513
Db 178 aAlaArgAlaProAspThrAlaAsnLeuThrThrAlaAlaLeuPheArgTrpArgThrAs 198
QY 514 CCGGCGCTGTGTCTGACCTGCTGACATCATCCCAAGATGSA----- 562
Db 198 nIleValIleCysSerProProLeuCysProGlyAlaProArgAlaArgAlaGluValGlu 218
QY 563 -ATCATCCCCGGGCTCAGTGGAGGCTCTGTCGATGCTCAACAGCCCTTTCAGTGGTGG 621
Db 218 yLeuValGluAspThrAlaCysGluArgSerCysProProValAlaIleArgAlaLeuGlu 238
QY 622 GAAGCGCTCGCGCTCTACAGCCAGCGGTGCTCAAGGAGTGGCTTCTCAACCTGCA 681
Db 238 uAlaGluCysArgLysTyrProValValIleLysAspValArgLeuLeuAspLeuGlu 258
QY 682 GTCCCTCTTACCCGCTGTGAAGACCCCTCCCTCACTCATATCTGTCACCTGCTCCG 741
Db 258 yValLeuValProLeuLeuArgAspProGlyLeuAsnLeuLysValValGlnLeuPheAr 278
QY 742 CGACCCCGCGCGCTGCTCCGTTCCCGAGAACGACAAAGGAGATCTCATGTTGTCAG 801
Db 278 sAspProArgAlaValHisAsnSerArgLeuLysSerArgGlnGlyLeuLeuArgGluSe 298
QY 802 T-----CGATTGTGATG----- 814
Db 298 rIleGlnValLeuArgThrArgGlnArgGlyAspArgPheHisArgValLeuLeuAlaHi 318
QY 815 -----GGGAGCATGAGCAAAACTCAAGAGGAGGAGGACCAACCCCTA 855
Db 318 eGlyValGlyAlaArgProGlyGlnSerArgAlaLeuProAlaAlaProArgAlaAs 338
QY 856 CTAT-----GTGATGCGAGTCTATGCTCCAAAGCCAGCGATGAGATCTTACAAGAC 903
Db 338 pPhePheLeuThrGlyAlaLeuGluValIleCysGluAlaTrpLeuArgAspLeuLeuPh 358
QY 904 CATCCAGCTCTTGGCCCAAGGCCCTGCGAGAACGCTACTGCTTGTGCGCTATGAGGACCT 963
Db 358 eAlaArgGlyAlaProAlaTrpLeuArgArgArgTyrLeuArgLeuArgTyrGluAspLe 378
QY 964 GGCTCGAGCCCTGTGGCCGACACTTCCCGAATGTATGAATTCGTGGATTTGGAATCTT 1023
Db 378 uValArgGlnProArgAlaGlnLeuArgArgLeuLeuArgPheSerGlyLeuArgAlaLe 398
QY 1024 GCCCATCTTCAGACCTGGTGCATACATCACCGAGGCAAGGCGCATGGT---GACCA 1080
Db 398 uAlaAlaLeuAspAlaPheAlaLeuAsnMetThrArgGlyAlaAlaTyrGlyAlaAspAr 418
QY 1081 CGCTTTCACACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTTGGCGCTGCTCTTT 1140
Db 418 gProPheHisLeuSerAlaArgAspAlaArgGluAlaValHisAlaTrpArgGluArgLe 438
QY 1141 GCCCTATGAAAAGTTCCTGACITTCAGAAAGCCCTGTGGCGATGCCATGAATTTGCTGG 1200
Db 438 uSerArgGluGlnValArgGlnValGluValAlaCysAlaProAlaMetArgLeuLeuAl 458
QY 1201 CTACGCCACGTCAGATCTGAACAGAA 1228
Db 458 aTyr-----ProArgSerGlyGluGlu 465

RESULT 13
O88276 PRELIMINARY; PRT; 530 AA.
ID O88276
AC O88276;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)

248	QY	GAGCCATGCACGCTGTCTCTCTCCTGGCGCTCTGCTCTTCTTTTGTCGGCGAG	307
163	Db	LysArgGlnLeuValTy ^r ValPheThrTrpArgSerGlySerPhePheGlyGlu	182
308	QY	CTTTTGGGCAGCACCCAGATGTTTCTACTCATGGAGCCGCCCTGGACAGCTGGGATG	367
183	Db	LeuPheAsnGlnAenProGluValPhePheLeuTy ^r GluProValTrpHisValTrpGln	202
358	QY	ACCITCAGCAGACGACCGCTCGATGCTGCACATGCTGTGGGATCATGATCGGGCC	427
203	Db	LysLeuTy ^r ProGlyAspAlaValserLeuGlnGlyAlaalaargaspMetIeuserAla	222
428	QY	GTTCTTTGTGCGACATGAGCGTCTTTGATGCCTTACATGCAACTCGTCCCCGGACAG	487
223	Db	LeuTy ^r ArgCysAspLeuSerValPheGlnLeuTy ^r SerProAladglseryGlyArg	242
488	QY	TCC-----AGCCTCTTTCAGTGGGAGACAGCGCGGCCCTGTGTCGCACCT	535
243	Db	AsnLeuThrThrLeuGlyllePheGlyAlaAlaThrAsnLysValValCysSerPro	262
536	QY	GCGCTGTGATCATATCCCACAAGATGAATCATATCCCCGGGCTCAGTCAGGTCCTG	592
263	Db	CysProAlaTy ^r --ArgLysGluValValGlyLeuValAspAspargValCyslys	281
593	QY	---TGCAGTCAACAGCCCTTTGAGTGGTGGAGAAGCCCTGCCGCTCTACAGCCACGTG	649
282	Db	LysCysProProGlnArgLeuAlaArgPheGlnGlnGluCysArgLysTy ^r ArgThrLeu	301
650	QY	GTGCTCAAGAGGTGCGCTTTCTCAACCTCAGTCCCTCTACCCGCTGTGTAAGACCC	709
302	Db	ValIleLysGlyValArgValPheAspValAlaValLeuAlaProLeuLysLysaspPro	321
710	QY	TCCCTCAACCTGCATATCGTCACCTGGTCGGGACCCCGCGCGCTGTTCGTTCCCGA	769
322	Db	AlaLeuaspLeuLysvalIleHisLeuValArgaspProargAlavalalaserSerArg	341
770	QY	GAACGCACAAAGGAGATCTCATGATTGACAGTCGCATTGTGTG-----	814
342	Db	IleArgSerArgHisGlyLeuIleArgGluSerLeuGlnValValArgSerargaspPro	361
815	QY	-----GGGCAGCATGAGCAAAAATCAGAGAGGAG	844
362	Db	ArgAlaHisArgMetProPheLeuGluAlaAlaGlyHisLysLeuGlyAlaLysLysGlu	381
845	QY	GACCAA-----CCCTACTAT-----GTGATGAGGTCTATCTGCCAACG	883

Db	382	GlyMetGlyGlyProAlaAspTyrHisAlaLeuGlyAlaMetGluValIleCysAsnSer	401
QY	884	CAGCTGGAGATCTACAGACCACTCCAGTCCTTGGCCCAAGGCCCTGCAGGAACGCTACCTG	943
Db	402	MetAlaIysThrLeuGlnThrAlaLeuGlnProProAspTyrLeuGlnGlyHisTyrLeu	421
QY	944	CTTGTGGCGCTATGAGGACCTGGCTGAGCCCTGTGGCCACAGCTTCCCGAATGTATGAA	1003
Db	422	ValValIleGlyTyrGluAspLeuValGlyAspProValIleThrLeuArgArgValTyrAsp	441
QY	1004	TTCGTGGGATGGAAATCTTTCGCCCATCTTCAGACCTGGGTGCATACATCACCGGAGGC	1063
Db	442	PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly	461
QY	1064	AAGGCGCATGGGTGACCAACGCTTCCACACAAATGCGAGGATGCCCTTAATGTCTCCACG	1123
Db	462	SerGlySerSerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn	481
QY	1124	GCTTGGCGCTGGTCTTTCGCCCATGAAAGTTTCTTCGACTTCAGAAAGCTGTGGCGAT	1183
Db	482	AlaIrpArgThrAlaLeuThrPheGlnGlnIleLysGlnValGluGluPheCysTyrGln	501
QY	1184	GCCATGAATTTGTGGGCTACCCGCCACGTCAGATCTGAACAGAACACAGAAACCTGTG	1243
Db	502	ProMetAlaValLeuGlyTyrGluArgValAsnSerProGluGluValIleAspLeuSer	521
QY	1244	CTGGATCTTCTG	1255
Db	522	LysThrLeuLeu	525
QY	1255	CTGGATCTTCTG	1265
Db	522	LysThrLeuLeu	525
QY	1265	CTGGATCTTCTG	1275
Db	522	LysThrLeuLeu	525
QY	1275	CTGGATCTTCTG	1285
Db	522	LysThrLeuLeu	525
QY	1285	CTGGATCTTCTG	1295
Db	522	LysThrLeuLeu	525
QY	1295	CTGGATCTTCTG	1305
Db	522	LysThrLeuLeu	525
QY	1305	CTGGATCTTCTG	1315
Db	522	LysThrLeuLeu	525
QY	1315	CTGGATCTTCTG	1325
Db	522	LysThrLeuLeu	525
QY	1325	CTGGATCTTCTG	1335
Db	522	LysThrLeuLeu	525
QY	1335	CTGGATCTTCTG	1345
Db	522	LysThrLeuLeu	525
QY	1345	CTGGATCTTCTG	1355
Db	522	LysThrLeuLeu	525
QY	1355	CTGGATCTTCTG	1365
Db	522	LysThrLeuLeu	525
QY	1365	CTGGATCTTCTG	1375
Db	522	LysThrLeuLeu	525
QY	1375	CTGGATCTTCTG	1385
Db	522	LysThrLeuLeu	525
QY	1385	CTGGATCTTCTG	1395
Db	522	LysThrLeuLeu	525
QY	1395	CTGGATCTTCTG	1405
Db	522	LysThrLeuLeu	525
QY	1405	CTGGATCTTCTG	1415
Db	522	LysThrLeuLeu	525
QY	1415	CTGGATCTTCTG	1425
Db	522	LysThrLeuLeu	525
QY	1425	CTGGATCTTCTG	1435
Db	522	LysThrLeuLeu	525
QY	1435	CTGGATCTTCTG	1445
Db	522	LysThrLeuLeu	525
QY	1445	CTGGATCTTCTG	1455
Db	522	LysThrLeuLeu	525
QY	1455	CTGGATCTTCTG	1465
Db	522	LysThrLeuLeu	525
QY	1465	CTGGATCTTCTG	1475
Db	522	LysThrLeuLeu	525
QY	1475	CTGGATCTTCTG	1485
Db	522	LysThrLeuLeu	525
QY	1485	CTGGATCTTCTG	1495
Db	522	LysThrLeuLeu	525
QY	1495	CTGGATCTTCTG	1505
Db	522	LysThrLeuLeu	525
QY	1505	CTGGATCTTCTG	1515
Db	522	LysThrLeuLeu	525
QY	1515	CTGGATCTTCTG	1525
Db	522	LysThrLeuLeu	525
QY	1525	CTGGATCTTCTG	1535
Db	522	LysThrLeuLeu	525
QY	1535	CTGGATCTTCTG	1545
Db	522	LysThrLeuLeu	525
QY	1545	CTGGATCTTCTG	1555
Db	522	LysThrLeuLeu	525
QY	1555	CTGGATCTTCTG	1565
Db	522	LysThrLeuLeu	525
QY	1565	CTGGATCTTCTG	1575
Db	522	LysThrLeuLeu	525
QY	1575	CTGGATCTTCTG	1585
Db	522	LysThrLeuLeu	525
QY	1585	CTGGATCTTCTG	1595
Db	522	LysThrLeuLeu	525
QY	1595	CTGGATCTTCTG	1605
Db	522	LysThrLeuLeu	525
QY	1605	CTGGATCTTCTG	1615
Db	522	LysThrLeuLeu	525
QY	1615	CTGGATCTTCTG	1625
Db	522	LysThrLeuLeu	525
QY	1625	CTGGATCTTCTG	1635
Db	522	LysThrLeuLeu	525
QY	1635	CTGGATCTTCTG	164

Db

116 LysArgGlnLeuValTyrValPheThrThrTrrPargSerGlySerPhePheGlyGlu 135

308 CTTTITGGGAGACCCAGATGTTTCTACCTGATGGAGCCCGCTGGCACCTGTGGATG 367

136 LeuPheAsnGlnAsnProGluValPhePheLeuTyrGluProValTrrPHisValTrrPglN 155

368 ACCTTCAAGCAGAGCACCCTGGATGCTGCATCGCTGTGGGATCTGTATACGGGCC 427

156 LysLeuTyrProGlyAspAlaValSerLeuGlnGlyAlaAlaArgAspMetLeuSerAla 175

428 GTCTTCTTGTGCACATGAGCGTCTTTGATGCTCATGGAACCTGGTCCCCGGAGACAG 487

176 LeuTyrArgCysAspLeuSerValPheGlnLeuTyrSerProAlaGlySerGlyGlyArg 195

488 TCC-----AGCTCTTTTCAGTGGGAGACAGCCGGCCCTGTGTTCTGCACCT 535

196 AsnLeuThrThrLeuGlyIlePheGlyAlaAlaThrAsnLysValValCysSerSerPro 215

536 GCCTGTGACATCATCCCAAGATGAATCATCCCCCGGGCTCACTGCAGGCTCCTG--- 592

216 LeuCysProAlaTyr--ArgLysGluValValGlyLeuValAspArgValCysLys 234

593 ---TGACGTCAACAGCCCTTGTAGTGTGGAGAGCCCTGCGCTCTACAGCCACGCTG 649

235 LysCysProGlnArgLeuAlaArgPheGluGluGluCysArgLysTyrArgThrLeu 254

650 GTCTCAAGGAGGTGGCTTTTCAACCTGCAGTCCCTCTACCCGCTGTGAAAGACCCC 709

255 ValIleLysGlyValArgValPheAspValAlaValAlaProLeuLeuArgAspPro 274

710 TCCTCTAACCTGCATATCGTGACCTGTGTCGGGACCCCGGGCCGTGTTCCTCCCGA 769

275 AlaLeuAspLeuLysValIleHisLeuValArgAspProArgAlaValAlaSerSerArg 294

770 GAACGCACAAAGGAGATCTCATGATTGACAGTCGCATTGTGATG----- 814

295 IleArgSerArgHisGlyLeuIleArgGluSerLeuGlnValValArgSerArgAspPro 314

815 -----GGCGACATGACGCAAAACTCAAGAAAGGAG 844

315 ArgAlaHisArgMetProPheLeuGluAlaAlaGlyHisLysLeuGlyAlaLysLysGlu 334

845 -----GACCAACCTACTAT-----GTGATGAGGTCTATCTGCCAAAGC 883

335 GlyValGlyGlyProAlaAspTyrHisAlaLeuGlyAlaMetGluValIleCysAsnSer 354

884 CAGCTGGAGATCTACAGACCATCCAGTCCCTGCCCAAGGCCCTGCAGGAACGCTACCTG 943

355 MetAlaLysThrLeuGlnThrAlaLeuGlnProProAspTrpLeuGlnGlyHisTyrLeu 374

944 CTTGTGGCTATGAGGACCTGGCTCGAGCCCTGTGCCCCAGACTTCCCGAATGTATGAA 1003

375 ValValArgTyrGluAspLeuValGlyAspProValLysThrLeuArgArgValTyrAsp 394

1004 TTCGTGGGATTGGAATTCCTTCCCTCATCTTTCAGACCTGGGTGCATACATCACCCGAGGC 1063

395 PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly 414

1064 AAGGCATGGTGACCGCTTTCACAAATCCAGGATGCCCTTAATGTCTCCCGAG 1123

415 SerGlySerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn 434

1124 GCTTGGCGCTGTCTTTCCTCTGAAAGGTTTCTCGACTCAGAAAGCCTGTGGCGAT 1183

435 AlaTrrPargThrAlaLeuThrPheGlnGlnIleLysGlnValGluGluPheCysTyrGln 454

1184 GCCATGAATTCCTGGGCTACCGCCAGCTCAGATCTGAACAGACAGAAACCTGTTG 1243

455 ProMetAlaValLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer 474

1244 CTGGATCTTCTG 1255

Db

475 LysThrLeuLeu 478

Search completed: May 6, 2004, 11:05:46

Job time : 124 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_r2p model

Run on: May 6, 2004, 10:52:03 ; Search time 96.5 seconds
(without alignments)
11963.618 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattccattgttgggta.....tgggattccataaaaaaaaaa 2043

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xlp
-Q/cgn2 1/USPTO.spool -P/US09645078/runat 06052004 104558 21717/app query.fasta_1.2183
-DB=A Geneseq 29Jan04 -QPMT=fastan -SUFFIX=tag -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR VAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09645078 @CGN 1.1.B1@runat 06052004 104558 21717 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq 29Jan04:
1: Geneseqp1980s:
2: Geneseqp1990s:
3: Geneseqp2000s:
4: Geneseqp2001s:
5: Geneseqp2002s:
6: Geneseqp2003as:
7: Geneseqp2003bs:
8: Geneseqp2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	54.5	386	2	AAY39918 Human gly
2	2035	54.4	386	2	AAM93309 Human pol
3	2008	53.7	380	5	Aau11274 Human L-s
4	1984	53.0	386	3	AAY79219 Human tra
5	1519.5	40.6	388	2	AAY39919 Mouse gly
6	1208.5	32.3	418	3	AAB41947 Human ORF
7	1030.5	27.5	418	5	ABB81557 Mouse int
8	1028.5	27.5	395	4	AAY72640 Human gly
9	1028.5	27.5	395	5	ABB81554 Human cor
10	1028.5	27.5	395	5	AAE15438 Human dru

11	1020	27.3	395	4	AAY72638	Aay72638 Mouse gly
12	1020	27.3	395	5	AAU11275	Aau11275 Murine in
13	1017.5	27.2	395	5	ABB81555	Abb81555 Consensus
14	1009	27.0	390	5	AAY72639	Aay72639 Human gly
15	1009	27.0	390	5	ABB81556	Abb81556 Human int
16	821	21.9	171	5	ABB81560	Abb81560 Human hig
17	614.5	16.4	486	6	ABP56121	Abp56121 Human cho
18	598.5	16.0	483	2	AAY31656	Aay31656 Mouse N-a
19	591.5	15.8	530	4	AAB95367	Aab95367 Human pro
20	587.5	15.7	484	2	AAY31657	Aay31657 Human N-a
21	587.5	15.7	531	5	AAU69414	Aau69414 Lung smal
22	549	14.7	411	2	AAW61100	Aaw61100 Keratan s
23	549	14.7	411	5	AAE25356	Aae25356 Human cho
24	549	14.7	411	6	ABU03503	Abu03503 Angiogene
25	519.5	13.9	169	5	ABB81558	Abb81558 Human cor
26	511.5	13.7	169	5	ABB81559	Abb81559 Human int
27	500.5	13.4	458	2	AAW06480	Aaw06480 Chick cho
28	492	13.1	499	6	ABR41139	Abr41139 Human DIT
29	482	12.9	479	2	AAW52863	Aaw52863 Glycosami
30	336	9.0	169	5	ABB81561	Abb81561 Human N-a
31	335.5	9.0	179	5	ABB81562	Abb81562 Human ker
32	293	7.8	174	5	ABB81563	Abb81563 Human cho
33	215	5.7	363	4	ABB64512	Abb64512 Drosophil
34	205	5.5	1212	2	AAW87503	Aaw87503 Human N-m
35	198.5	5.3	1061	2	AAW87504	Aaw87504 Human N-m
36	198.5	5.3	1938	6	ABB98398	Abb98398 Streptomy
37	198	5.3	1232	3	AAE28239	Aae28239 Human N-m
38	198	5.3	1232	5	AAW47961	Aaw47961 Human NMD
39	198	5.3	1232	6	ABU61439	Abu61439 Human N-m
40	198	5.3	1232	7	AAE39250	Aae39250 Human NMD
41	192.5	5.1	1081	3	AAE26240	Aae26240 Human N-m
42	192.5	5.1	1081	5	ABB56523	Abb56523 Human NMD
43	192.5	5.1	1081	6	ABU61440	Abu61440 Human N-m
44	192.5	5.1	1081	6	AAE39251	Aae39251 Human NMD
45	187.5	5.1	19938	6	ABB98398	Abb98398 Streptomy

ALIGNMENTS

RESULT 1
AAY39918
ID AAY39918 standard; protein; 386 AA.
XX
AC AAY39918;

XX 08-DEC-1999 (first entry)

XX Human glycosyl sulfotransferase-3 protein sequence.

XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
XX selectin binding interaction; inflammation; lymphocyte homing; human;
XX secondary lymph organ.

XX Homo sapiens.

XX WO9949018-A1.

XX 30-SEP-1999.

XX 26-FEB-1999; 99WO-US004316.

XX 20-MAR-1998; 98US-00045284.

XX 12-NOV-1998; 98US-00190911.

XX (REGC) UNIV CALIFORNIA.

XX (SYNT) SYNTEX USA INC.

XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;

XX WPI; 1999-580442/49.

XX N-PSDB; AAZ20792.

XX Human and murine glycosyl sulfotransferase 3 and related polynucleotides.

Db	241	GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle	260
QY	908	CAGTCCTTGGCCAGAGCCCTGCAGGAACGCTACCTGCTTGTCGGCTATGAGGACCTGGCT	967
Db	261	GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla	280
QY	968	CGAGCCCTGTGGCCACAGACTCCCGAATGTATGAATTCGTGGGATTCGGAATTCCTTGGCCC	1027
Db	281	ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro	300
QY	1028	CATCTTCAGACTGGGTGCATACATCACCCGAGGCAGGGCATGGGTGACACAGCCTTC	1087
Db	301	HisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe	320
QY	1088	CACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTGGCGCTGGTCTTTCGCCCTAT	1147
Db	321	HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTyrSerLeuProTyr	340
QY	1148	GAAGAAGTTTCTCGACATTCAGAAAGCCTGTGGCGATGCCATGAATTCCTGGGTACCGC	1207
Db	341	GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrArg	360
QY	1208	CAGCTCAGATCTGAACAGAACAGAGAACCCTGTGTGCTGGATCTTCTCTACTCTGAGCT	1267
Db	361	HisValArgSerGluGlnGluArgAsnLeuLeuLeuAspLeuSerThrTrpThr	380
QY	1268	GTCCCTGAGCAAAATCCAC	1285
Db	381	ValProGluGlnIleHis	386

1088	CACACAAATGCCAGGATGCCCTTAATGCTCCCAAGCTTGCGCTGCTTTCGCCCTAT	1147
QY		
321	HsThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTyr	340
Db		
1148	GAAGAAGTTTCTCGACTTCAGAAGCCTGTGGCGATGCCATGAATTCCTGGGTACGCG	1207
QY		
341	GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuGlyTyrArg	360
Db		
1208	CACCTCAGATCTGACACAGAACAGAGAACCCTGTTGCTGGATCTTCCTACCTGGACT	1267
QY		
361	HsValArgSerGluGlnGluGlnArgAsnLeuLeuAspLeuLeuSerThrTrpThr	380
Db		
1268	GTCCCTGAGCAATCCAC	1285
QY		
381	ValProGluGlnIleHis	386
Db		

[illegible]

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WFI; 2001-524355/58.
DR N-PSDB; AAK94229.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in generic manipulation.
PT
XX Claim 8; SEQ ID NO 2817; 1380pp + Sequence Listing; English.
PS
XX The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesising the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full

CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a polypeptide encoded by a full length
 CC human cDNA of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in CD-ROM
 CC format directly from EPO

XX Sequence 386 AA;

Alignment Scores:

Pred. No.: 1,92e-172 Length: 386
 Score: 2035.00 Matches: 385
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.74% Mismatches: 0
 Query Match: 54.38% Indels: 0
 DB: 4 Gaps: 0

US-09-645-078-1 (1-2043) x AM93309 (1-386)

QY 128 ATGCTACTGCTAAATAAATGAAGTCTCTGCTGTTCTGTTTCCAGATGGCCATCTTG 187
 Db 1 MetLeuLeuProLysLysMetLysLeuLeuPheLeuValSerGlnMetAlaIleLeu 20
 QY 188 GCTCTATTCTCCACATGATGACGACACACATCAGCTCCCTGTCTATGAGGACACAGCC 247
 Db 21 AlaLeuPhePheHisMetTy-SerHisAsnIleSerSerLeuSerMetLysAlaGlnPro 40
 QY 248 GAGCCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307
 Db 41 GluArgMetHisValLeuValLeuSerSerIrrpArgSerGlySerPheValGlyGln 60
 QY 308 CTTTGTGGGAGCACCAGATGTTTCTACCTGATGAGCCCGCTGCGCATGCTGATACGGGCC 427
 Db 81 ThrPheLysGlnSerThAlaTrpMetLeuHisMetAlaValArgAspLeuIleArgAla 100
 QY 428 GTCTTCTTGTGCGATGAGCGTCTTCTGATGCTTACATGAGAACCTGTGTCCTGAGGACAG 487
 Db 101 ValPheLeuLysAspMetSerValPheAspAlaTyMetGluProCgIyProArgArgGln 120
 QY 488 TCCAGCCTCTTTCAGTGGGAGACAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 547
 Db 121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuLysSerAlaProAlaCysAspIle 140
 QY 548 ATCCCAAGATGAATCATCCCCGGGCTCACTGACGGCTCTGTCAGTCAACAGGCC 607
 Db 141 IleProGlnAspGluIleIleProArgAlaHisCysArgLeuLeuLysSerGlnGlnPro 160
 QY 608 TTTGAGTGTGGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 667
 Db 161 PheGluValValGluLysAlaCysArgSerTySerHisValValLeuLysGluValArg 180
 QY 668 TTCTTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
 Db 181 PhePheAsnLeuGlnSerLeuTyProLeuLeuLysAspProSerLeuAsnLeuHisIle 200
 QY 728 GTGACCTGTGTCGGGAGACCCCGGCGGTGTCCTCCGAGAGACGACAAAGGAGAT 787
 Db 201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAsp 220
 QY 788 CTCATGATTGACGTGCGATTTGATGGGCGACATGAGCAAACTCAAGAGGAGGAC 847
 Db 221 LeuMetIleAspAsnArgIleValMetGlyGlnHisGluGlnLysLeuLysGluAsp 240
 QY 848 CAACCTTACTATGTGATGAGGTGATCTGTCGCAAGCCAGCTGGAGATCTACAGACCATC 907
 Db 241 GlnProTyTyValMetGlnValIleCysGlnSerGlnLeuGluIleTyLysThrIle 260
 QY 908 CAGTCTCTTCCCAAGCCCTGAGGAAACGCTACCTGCTGTGCTGCTGCTGCTGCTGCTGCT 967

Db 261 GlnSerLeuProLysAlaLeuGlnGluArgTyLeuLeuValArgTyLeuAspLeuAla 280
 QY 968 CGAGCCCTGTGGCCGACATCCCGAATGATGATGATGATGATGATGATGATGATGATGATG 1027
 Db 281 ArgAlaProValAlaGlnThrSerArgMetTyGluPheValGlyLeuGluPheLeuPro 300
 QY 1028 CATCTTCAGACCTGGGTGCATACATCACCCGAGGCAAGGCGATGGGTGACACGCTTTC 1087
 Db 301 HisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
 QY 1088 CACACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTTGGCGTGTCTTGTCCCTAT 1147
 Db 321 HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTy 340
 QY 1148 GAAAAGGTTTCTCGACTTCAGAAAGCCTGTGCGCATGCGCATGAAATTTCTGCGGTACCG 1207
 Db 341 GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyArg 360
 QY 1208 CAGCTCAGATCTGAACAAGACAGAGAAACCTGTGTGCTGCTGCTGCTGCTGCTGCTG 1267
 Db 361 HisValArgSerGluGlnGlnGlnArgAsnLeuLeuLeuLeuLeuLeuSerThrTrpThr 380
 QY 1268 GTCCCTGAGCAAAATCCAC 1285
 Db 381 ValProGluGlnIleHis 386
 RESULT 3
 AAU11274
 ID AAU11274 standard; protein; 380 AA.
 XX AAU11274;
 XX 12-MAR-2002 (first entry)
 DE Human L-selectin sulfotransferase-2 (LSST-2) protein.
 XX Human; beta1,3GnT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic.
 XX Homo sapiens.
 XX WO200185177-A1.
 XX 15-NOV-2001.
 PF 10-MAY-2001; 2001WO-US015452.
 XX 11-MAY-2000; 2000US-00569320.
 PR (BURN-) BURNHAM INST.
 PA Fukuda M, Yeh J, Hiraoka N;
 PI WPI: 2002-075226/10.
 DR N-PSDB; AAS16947.
 XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
 PT L-selectin sulfotransferase-2 that directs expression of L-selectin
 PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
 PT GlcNAc 6-sulfotransferase.
 XX Claim 21; Fig 4; 98pp; English.
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated beta1,3-N-
 CC acetylglucosaminyltransferase (beta1,3GnT) or an active fragment, where
 CC beta1,3GnT directs expression of a MECA-79 antigen. The invention also

Mouse glycosyl sulfotransferase-3 protein sequence.

Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing; selectin binding interaction; inflammation; lymphocyte homing; mouse; secondary lymph organ.

Mus sp.

WO9949018-A1.

30-SEP-1999.

26-FEB-1999; 99WO-US004316.

20-MAR-1998; 98US-00045284.

12-NOV-1998; 98US-00190311.

(REGC) UNIV CALIFORNIA.

(SYNT) SYNTAX USA INC.

Bistrup A, Rosen SD, Tangemann K, Hemmerich S;

WPI; 1999-580442/49.

N-PSDB; AAZ20793.

Human and murine glycosyl sulfotransferase 3 and related polynucleotides.

Claim 2; Fig 4; 59pp; English.

This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs

Sequence 388 AA;

Alignment Scores:

Pred. No.: Length: 388

Score: 1519.50 Matches: 283

Percent Similarity: 85.13% Conservative: 49

Best Local Similarity: 72.56% Mismatches: 55

Query Match: 40.61% Indels: 3

DB: Gaps: 2

US-09-645-078-1 (1-2043) x AAY39919 (1-388)

QY 128 ATGCTACTGCCTAATAAATGAAGCTCTGCTGTTCTTCAGATGCCATCTTG 187

Db 1 MetMetLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuVal 20

QY 188 GCTCTATTCTCCACATGTACAGCCACACATCATGCTCCCTGTCTATGAAGGCACAGCC 247

Db 21 AlaLeuPheIleHisMetSerValHisArg-----HisLeuSerGlnArgGluGluSer 38

QY 248 GAGCGC---ATCAGCGTCTGGTCTGCTTCCTCTGGCGCTCTGGCTCTCTTTTGGGG 304

Db 39 A-GA-TG-ProValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGly 58

QY 305 CAGCTTTTGGGCAGCACCAGATGTTTCTACCTGATGAGCGCCCGCTGGACGCTGC 364

Db 59 GlnLeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrp 78

QY 365 ATGACCTTCAAGCAGACGACCGCTGGATGCTGCACATGGCTGTGGGGATCTGATACGG 424

Db 79 MetThrPheThrSerSerThrAlaTrpLysLeuHisMetAlaValArgAspLeuLeuArg 98

QY 425 CGCGTCTCTTGTGGCAGATGAGCGCTCTTGATCGCTACATGGAACTGCTCCCGGAGA 484

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; Gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
XX
XX 02-APR-1999; 99US-0127636P.
XX
XX 05-APR-1999; 99US-0127728P.
XX
XX 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX
XX N-PSDB; AAC76156.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 2599-2600; 5507pp; English.
XX
XX AAC74446 to AAC7606 encode the proteins given in ABA40237 to ABA43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX Sequence 418 AB;

Alignment Scores:
Pred. No.: 1.17e-98 Length: 418
Score: 1208.50 Matches: 249
Percent Similarity: 69.34% Conservative: 54
Best Local Similarity: 56.98% Mismatches: 85
Query Match: 32.30% Indels: 49
DB: 3 Gaps: 10

US-09-645-078-1 (1-2043) x ABA41947 (1-418)

QY 29 GACAGAGGAGGTAGAGAGAAAGCGCA-----TGGCCGGCTAGCAGTGCCTCT 79
|||:|||||
Db 10 AspGlnArgCysSerSerProSerSerProGlyArgTyrPro----- 23

QY 80 CAAAGACAGCAGGAGAGCCCAAGCCCAAGGCTTC-----CACTTCAGCACA----- 127
|||:|||||
Db 24 -----ProAlaAlaArgMetTyrLeuProArgPheSerSerIysThr 37
|||:|||||
QY 128 -----ATCCTACTGCTTAAAAAATGAAGCTCTGCTGTTCTGTTTCCAGATG 178
|||:|||||
Db 38 ValThrValLeuLeuAlaGlnThrThrCysLeuLeuLeuPheIleIleSerArgPro 57
|||:|||||
QY 179 GCATCTTGGCTCTATTCTTCCACATGTACAGCCACAAACATCAGCTCCCTGTTCTGAAG 238
|||:|||||
Db 58 -----GlyProSerSerProAla 63
|||:|||||
QY 239 GCACAGCCGAGGAGGATGACGTGCTGTTCTTCTCTGCGCTGCTGCTTCTTTT 298
|||:|||||
Db 64 GlyGlyGluAspArgValHisValLeuValLeuSerSerTyrArgSerGlySerPhe 83
|||:|||||
QY 299 GTGGGCGAGCTTTTGGGCGAGCAGCAGATGTTTCTTACCTCATGAGAGCCCGCTGGCAC 358
|||:|||||
Db 84 LeuGlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTyrHis 103
|||:|||||
QY 359 GTGTGGATGACCTTCAAGCAGAGCAGCCGCTGGATGTGCATGCTGTGGGATCTG 418
|||:|||||
Db 104 ValTyrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeu 123
|||:|||||
QY 419 ATAGGGCGCTCTTGTGTGCGACATGAGCGCTTTGTATGCTACATGAGAACTGGTCC 478
|||:|||||
Db 124 MetArgSerIlePheLeuCysAspMetAspValPheAspAlaTyrMetGluProGlyPro 143
|||:|||||
QY 479 CGGAGACAGTCCAGCCTCTTTCAGTGGGAGAACAGCCGGCGCTGTGTTCTGCACCTGCC 538
|||:|||||
Db 144 ArgArgGlnSerSerLeuPheGlnTyrGluAsnSerArgAlaLeuCysSerAlaProAla 163
|||:|||||
QY 539 TGTGACATATCCACAGATGAATCATCCCGGGCTCACTGAGGCTCTGTGTCAGT 598
|||:|||||
Db 164 CysAspIleLeuProGlnAspGluIleIleProAlaGlnHisCysArgLeuLeuCysSer 183
|||:|||||
QY 599 CAACAGCCCTTTGAGGTGTGGAGAGGCTCCCGCTCTACAGCCACAGTGTGTCTCAAG 658
|||:|||||
Db 184 GlnGlnProPheGluValValGluLysAlaCysArgSerTyrSerHisValValLeuLys 203
|||:|||||
QY 659 GAGTGGCTTCTTCAACCTGAGTCTCTACCGCTCTGCTGAAAGACCCCTCCCTCAC 718
|||:|||||
Db 204 GluValArgPhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsn 223
|||:|||||
QY 719 CTGATATCTGTGACCTGTGTCGGGACCCCGGGCGCTGTTCCTCCGAGAGAGCACA 778
|||:|||||
Db 224 LeuHisIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluAlaAla 243
|||:|||||
QY 779 AAGCGAGATCTCATGATTGACATGTCATGATGGGCGCAGCATGAGCAAACTCAAG 838
|||:|||||
Db 244 GlyProIleLeuAlaArgAspAsnGlyIleValLeuGlyThrAsn---GlyLysTrpVal 262
|||:|||||
QY 839 AAGGAGAGCAACCTACTATGTGTCAGGTCATCTGCCAAAGCCAGCTGGAGATCTAC 898
|||:|||||
Db 263 GluAlaAspProHisLeuArgLeuIleIleArgGluValCysArgSerHisValArgIleAla 282
|||:|||||
QY 899 AAG-----ACCATCCAGTCTTTCGAGGAGGCGCTGAGAGAGCGTCTGTTGTGCGC 952
|||:|||||
Db 283 GluAlaAlaThrLeuLysPro---ProProPheLeuArgGlyArgTyrArgLeuValArg 301
|||:|||||
QY 953 TATGAGGACCTGCTGAGCCCTGTCGGCCGACACTTCCGATGTATCAATTCGTGGGA 1012
|||:|||||
Db 302 PheGluAspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThrGly 321
|||:|||||
QY 1013 TTGGAATCTTGGCCCATCTTTCAGACCTGGTGTGCATTAACATCACCCGAGGAGGCGATG 1072
|||:|||||
Db 322 LeuThrLeuThrProGlnLeuGluAlaTyrIleHisAsnIleThrHisGlySerGlyIle 341
|||:|||||
QY 1073 GTTGAC-----CAGCTTTCCACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCT 1126
|||:|||||
Db 342 GlyLysProIleGluAlaPheHisThrSerSerArgAsnAlaArgAsnValSerGlnAla 361
|||:|||||
QY 1127 TGGCGCTGGTCTTTGGCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCTGTGCGCATGCC 1186
|||:|||||

Db 362 TrpArgHisAlaLeuProPheThrLysileLeuArgValGlnGluValCysAlaGlyAla 381
 QY 1197 ATGAATTGCTGGCTACCGCCAGTCAGATCTGAACAAGACAGAGAACTGTTGCTG 1246
 Db 382 LeuGlnLeuGlyTyrArgProValTyrSerAlaaspGlnGlnArgaspLeuThrLeu 401
 QY 1247 GATCTTCG-----TCTACCTGGACTGTCCTCGAG 1276
 Db 402 AspLeuValLeuProArgGlyProAspHisPheSerTrpAlaSerProAsp 418
 RESULT 7
 ABB81557
 ID ABB81557 standard; protein; 418 AA.
 XX
 AC ABB81557;
 DT 05-SEP-2002 (first entry)
 XX Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
 DE Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 XX Mus musculus.
 OS
 XX US2002061562-A1.
 PN
 XX 23-MAY-2002.
 XX 09-AUG-2001; 2001US-00927602.
 XX 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.
 XX (FUKU/) FUKUDA M N.
 PA (AKAMA/) AKAMA T O.
 XX
 PI Fukuda MN, Akama TO;
 XX WPI; 2002-507643/54.
 XX
 PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX Example 5; Page 24-25; 69pp; English.
 XX
 CC The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratoplasty or keratectomy. The present sequence
 CC represents mouse intestinal N-acetylglucosamine-6- sulfotransferase,
 CC which is given in comparison with (I) in the exemplification of the
 CC present invention
 CC
 XX Sequence 418 AA;
 SQ

Alignment Scores:
 Pred. No.: 9,07e-83 Length: 418
 Score: 1030.50 Matches: 221
 Percent Similarity: 66.75% Conservative: 44
 Best Local Similarity: 55.67% Mismatches: 120
 Query Match: 27.54% Indels: 12
 DB: 5 Gaps: 7

US-09-645-078-1 (1-2043) x ABB81557 (1-418)

QY 128 ARGCTACTGCCTAA-----AAATGAGCTCCTGCTGTTCTTGGTTTCCAGATGCC 181
 Db 25 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuMetValGlnThrGly 44
 QY 182 ATCTTGGCTCTATTCTTCCACATGTACAGCCACACATCAGTCCTCTCTATGAGSCA 241
 Db 45 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 60
 QY 242 CAGCCCGAGCGCATCGCTGCTGTTCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 301
 Db 61 LeuGlyGluArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheVal 80
 QY 302 GGGCAGCTTTTGGCAGCACCAGATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 361
 Db 81 GlyGlnLeuPheSerGlnHisProaspValPheTyrLeuMetGluProAlaTrpHisVal 100
 QY 362 TGGATGACCTTCAAGCAGACACCGCTGGATGTGTCATCGCTGCTGCTGCTGCTGCTGATA 421
 Db 101 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeuIle 120
 QY 422 CGGGCGCTCTTCTTGGCAGATGAGCGTCTTGTATGCTTACATGAGAACCTGCTCCCGG 481
 Db 121 ArgSerValPheLeuCysaspMetaspValPheAspAlaTyrLeu---ProTrpArg 139
 QY 482 AGACACTCCAGCTCTTTCAGTGGAGAACACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGT 541
 Db 140 AsnIleSerAspLeuPheGlnTrpAlaValSerAlaLeuCysSerProProValCys 159
 QY 542 GACATATCCACAGATGAATCATCCCGGGCTCACTGAGGCTCCTGCTGCTGCTGCTGCTG 601
 Db 160 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLeuProLeuCysAlaThr 179
 QY 602 CAGCCCTTTGAGGTGGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661
 Db 180 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 199
 QY 662 GTGGCTTCTTCAACCTGAGTCCCTTACCGCTGCTGAGAGACCCCTCCCTCAACCTG 721
 Db 200 ValArgPheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 219
 QY 722 CATATGCTGACCTGCTCGGGACCCCGGGCGCTGTTCCGTTCCCGAGACGACCAAG 781
 Db 220 ArgIleValHisLeuValArgaspProArgAlaValLeuArgSerArgGlnGlnThrAla 239
 QY 782 GAGATCTCATGATTGACAGTCCATGTTGATGGGCGACATGAGCAAACTCAAGAG 841
 Db 240 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---Glu 258
 QY 842 GAGGACCAACCTACTATGTGTGTCAGGTCTCTGCCAAAGCCAGCTGAGATCTACAAG 901
 Db 259 AlaaspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 278
 QY 902 ACCATCCAGTCTTGGCCAGGCGCTGAGGAACGCTACTCTGCTTGTGCTGCTGCTGCTG 961
 Db 279 AlaLeuHisLysProProPheLeuGlnAspArgTyrArgLeuValArgTyrGluAsp 298
 QY 962 CTGCTCGAGCCCTGCTGCGCCAGACTTCCCGAATGTATGAATTCGTGGATTTGGAATTC 1021
 Db 299 LeuAlaArgaspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGlyLeu 318
 QY 1022 TTGCCCATCTTTCAGACTCGGTGTCATACATCACCGGAGGCAAGGGCATGGGT----- 1075
 Db 319 ThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAlaArg 338
 QY 1076 GACCAGCTTTCCACACAAATGCCAGGATGCCCTTAATGCTCTCCAGGGCTGCGCTGG 1135
 Db 339 ArgGluAlaPheLysThrThrSerArgaspAlaLeuSerValSerGlnAlaTrpArgHis 358
 QY 1136 TCTTTGGCTATGAAAGGTTTCTCGACTTCAGAAAGCTGTGCGGATGCGCATGATTTG 1195
 Db 359 ThrLeuProPheAlaLysIleArgValGlnGlnLeuCysGlyGlyAlaLeuGlnLeu 378
 QY 1196 CTGGGCTACCGCCAGCTCAGTCTGTAACAGACAGAGAACCTGTTGCTGGATCTTCTG 1255

Db 379 LeuGlyTyrArgSerValHisSerGluLeuGluGlnArgAspLeuSerLeuAspLeuLeu 398
QY 1256 TCTACCT---GGACTGTCCTCGAGCAAAATCCACTAAGAGGGGTGAGAG 1301
Db 399 -LeuProArgGlyMetAspSerPheLysTrpAlaSerSerThrGluLys 414
RESULT 8
AAY72640
ID AAY72640 standard; protein; 395 AA.
XX AC
XX AAY72640;
XX
DT 02-MAY-2001 (first entry)
DE Human glycosyl sulfotransferase-4beta (GST-4beta).
XX
KW Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1.
XX
XX Homo sapiens.
OS
FN WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US019741.
XX
XX 20-JUL-1999; 99US-0144694P.
PR
XX 13-JUN-2000; 2000US-00593828.
PR
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX
XX WPI; 2001-138471/14.
DR
XX N-PSDB; AAD02697, AAD02700.
DR
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications.
XX
XX Claim 3; Fig 4B; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
XX beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
XX membrane protein useful for inhibiting a binding event between a selectin
XX and a selectin ligand, which comprises contacting the selectin with a nor-
XX -sulphated selectin ligand, GST and a small molecular agent that inhibits
XX the sulphation activity of GST. GST is also useful in inhibiting a
XX selectin mediated binding event. GST is useful in gene therapy to treat
XX disorders such as acute or chronic inflammation, systemic lupus
XX erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
XX polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation
XX
XX Sequence 395 AA;
SQ

Alignment Scores:	1.34e-82	Length:	395
Pred. No.:			

Score:	1028.50	Matches:	224
Percent Similarity:	65.06%	Conservative:	46
Best Local Similarity:	53.98%	Mismatches:	108
Query Match:	27.49%	Indels:	37
DB:	4	Gaps:	10

US-09-645-078-1 (1-2043) x AAY72640 (1-395)

Qy	125	ACAATGCTACTGCTCTAAAAAATAAGACTCTGCTGTTTCTGGTTTCCAGATGGCCATC	184
Db <td>12<td>ThrAlaLeuLeuAlaGlnThrPheLeuLeuPheLeuValSerArgPro</td><td>29</td></td>	12 <td>ThrAlaLeuLeuAlaGlnThrPheLeuLeuPheLeuValSerArgPro</td> <td>29</td>	ThrAlaLeuLeuAlaGlnThrPheLeuLeuPheLeuValSerArgPro	29
Qy <td>185<td>TTGGCTCTATTCTCCACATGTACAGCCACAATCAGCTCCTGTCTTATGAAGGCACAG</td><td>244</td></td>	185 <td>TTGGCTCTATTCTCCACATGTACAGCCACAATCAGCTCCTGTCTTATGAAGGCACAG</td> <td>244</td>	TTGGCTCTATTCTCCACATGTACAGCCACAATCAGCTCCTGTCTTATGAAGGCACAG	244
Db <td>30</td> <td>-----GlyProSerSerProAlaGlyGly</td> <td>37</td>	30	-----GlyProSerSerProAlaGlyGly	37
Qy <td>245<td>CCGAGCGCATGCATGCTGTTGTTCTTCCTGGCGCTCTGGCTCTCTCTTTTGTGGGG</td><td>304</td></td>	245 <td>CCGAGCGCATGCATGCTGTTGTTCTTCCTGGCGCTCTGGCTCTCTCTTTTGTGGGG</td> <td>304</td>	CCGAGCGCATGCATGCTGTTGTTCTTCCTGGCGCTCTGGCTCTCTCTTTTGTGGGG	304
Db <td>38<td>GluAlaArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGly</td><td>57</td></td>	38 <td>GluAlaArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGly</td> <td>57</td>	GluAlaArgValHisValLeuValLeuSerSerTrpArgSerGlySerPheValGly	57
Qy <td>305<td>CACGTTTGGGAGCAGCACCCAGATGTTTCTACCTCATGAGGCCCGCCGAGCAGTGTCG</td><td>364</td></td>	305 <td>CACGTTTGGGAGCAGCACCCAGATGTTTCTACCTCATGAGGCCCGCCGAGCAGTGTCG</td> <td>364</td>	CACGTTTGGGAGCAGCACCCAGATGTTTCTACCTCATGAGGCCCGCCGAGCAGTGTCG	364
Db <td>58<td>GlnLeuPheAsnGlnHisProAspValPheTrpLeuMetGluProAlaTrpHisValTrp</td><td>77</td></td>	58 <td>GlnLeuPheAsnGlnHisProAspValPheTrpLeuMetGluProAlaTrpHisValTrp</td> <td>77</td>	GlnLeuPheAsnGlnHisProAspValPheTrpLeuMetGluProAlaTrpHisValTrp	77
Qy <td>365<td>ATGACCTTCAAGCAGACAGCCGCTCGATGCTGCACATGCTGTGCGGATCTGATACGG</td><td>424</td></td>	365 <td>ATGACCTTCAAGCAGACAGCCGCTCGATGCTGCACATGCTGTGCGGATCTGATACGG</td> <td>424</td>	ATGACCTTCAAGCAGACAGCCGCTCGATGCTGCACATGCTGTGCGGATCTGATACGG	424
Db <td>78<td>ThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuValArg</td><td>97</td></td>	78 <td>ThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuValArg</td> <td>97</td>	ThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuValArg	97
Qy <td>425<td>GCGTCTCTTTGTCGACATGACGCTCTTTGTATGCTCTACATGGAACTGTCCTCCCGAGA</td><td>484</td></td>	425 <td>GCGTCTCTTTGTCGACATGACGCTCTTTGTATGCTCTACATGGAACTGTCCTCCCGAGA</td> <td>484</td>	GCGTCTCTTTGTCGACATGACGCTCTTTGTATGCTCTACATGGAACTGTCCTCCCGAGA	484
Db <td>98<td>SerValPheLeuCysAspMetAspValPheAspAlaTrpLeu---ProTrpArgArgAsn</td><td>116</td></td>	98 <td>SerValPheLeuCysAspMetAspValPheAspAlaTrpLeu---ProTrpArgArgAsn</td> <td>116</td>	SerValPheLeuCysAspMetAspValPheAspAlaTrpLeu---ProTrpArgArgAsn	116
Qy <td>485<td>CAGTCCAGCTCTTTTCAGTGGGAGAACAGCCGGCGCTGTGTCTGCACCTGCTCTGAC</td><td>544</td></td>	485 <td>CAGTCCAGCTCTTTTCAGTGGGAGAACAGCCGGCGCTGTGTCTGCACCTGCTCTGAC</td> <td>544</td>	CAGTCCAGCTCTTTTCAGTGGGAGAACAGCCGGCGCTGTGTCTGCACCTGCTCTGAC	544
Db <td>117<td>LeuSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProAlaCysSer</td><td>136</td></td>	117 <td>LeuSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProAlaCysSer</td> <td>136</td>	LeuSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProProAlaCysSer	136
Qy <td>545<td>ATCATCCACAAGATGAATCATCCCGGGCTCACTGAGGCTCTGTGTCATCAACAG</td><td>604</td></td>	545 <td>ATCATCCACAAGATGAATCATCCCGGGCTCACTGAGGCTCTGTGTCATCAACAG</td> <td>604</td>	ATCATCCACAAGATGAATCATCCCGGGCTCACTGAGGCTCTGTGTCATCAACAG	604
Db <td>137<td>AlaPheProArgGlyAlaIleSerSerGluAlaValCysLysProLeuCysAlaArgGln</td><td>156</td></td>	137 <td>AlaPheProArgGlyAlaIleSerSerGluAlaValCysLysProLeuCysAlaArgGln</td> <td>156</td>	AlaPheProArgGlyAlaIleSerSerGluAlaValCysLysProLeuCysAlaArgGln	156
Qy <td>605<td>CCCTTGAGTGTGGAGAGGCTCCGCTCTCAGCCACCTGTGTCTCAAGGAGTG</td><td>664</td></td>	605 <td>CCCTTGAGTGTGGAGAGGCTCCGCTCTCAGCCACCTGTGTCTCAAGGAGTG</td> <td>664</td>	CCCTTGAGTGTGGAGAGGCTCCGCTCTCAGCCACCTGTGTCTCAAGGAGTG	664
Db <td>157<td>SerPheThrLeuAlaArgGluAlaCysArgSerTrpSerHisValValLeuLysGluVal</td><td>176</td></td>	157 <td>SerPheThrLeuAlaArgGluAlaCysArgSerTrpSerHisValValLeuLysGluVal</td> <td>176</td>	SerPheThrLeuAlaArgGluAlaCysArgSerTrpSerHisValValLeuLysGluVal	176
Qy <td>665<td>CGCTTCTCAACCTGCAGTCCCTCTACCGCTCTGAAAGACCCCTCCCTCAACCTGCAT</td><td>724</td></td>	665 <td>CGCTTCTCAACCTGCAGTCCCTCTACCGCTCTGAAAGACCCCTCCCTCAACCTGCAT</td> <td>724</td>	CGCTTCTCAACCTGCAGTCCCTCTACCGCTCTGAAAGACCCCTCCCTCAACCTGCAT	724
Db <td>177<td>ArgPhePheAsnLeuGlnValLeuTrpProLeuLeuSerAspProAlaLeuAsnLeuArg</td><td>196</td></td>	177 <td>ArgPhePheAsnLeuGlnValLeuTrpProLeuLeuSerAspProAlaLeuAsnLeuArg</td> <td>196</td>	ArgPhePheAsnLeuGlnValLeuTrpProLeuLeuSerAspProAlaLeuAsnLeuArg	196
Qy <td>725<td>ATCGTGACCTGTCGGGACCCCGGCGCTGTTCCGTTCCGAGAACCCACAAAGGA</td><td>784</td></td>	725 <td>ATCGTGACCTGTCGGGACCCCGGCGCTGTTCCGTTCCGAGAACCCACAAAGGA</td> <td>784</td>	ATCGTGACCTGTCGGGACCCCGGCGCTGTTCCGTTCCGAGAACCCACAAAGGA	784
Db <td>197<td>IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys</td><td>216</td></td>	197 <td>IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys</td> <td>216</td>	IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys	216
Qy <td>785<td>GATCTCATGTTGACAGTCGCATTTGATGGGCGCATGAGCAAAAACCTCAAGAAGGAG</td><td>844</td></td>	785 <td>GATCTCATGTTGACAGTCGCATTTGATGGGCGCATGAGCAAAAACCTCAAGAAGGAG</td> <td>844</td>	GATCTCATGTTGACAGTCGCATTTGATGGGCGCATGAGCAAAAACCTCAAGAAGGAG	844
Db <td>217<td>AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla</td><td>235</td></td>	217 <td>AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla</td> <td>235</td>	AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla	235
Qy <td>845<td>GACCAACCTCTATGTGATGCAAGTCATCTGCCAAGCAGCTGGAGATCTACAAG---</td><td>901</td></td>	845 <td>GACCAACCTCTATGTGATGCAAGTCATCTGCCAAGCAGCTGGAGATCTACAAG---</td> <td>901</td>	GACCAACCTCTATGTGATGCAAGTCATCTGCCAAGCAGCTGGAGATCTACAAG---	901
Db <td>236<td>AspProGlyLeuArgValAlaValGluValCysArgSerHisValArgIleAlaGluAla</td><td>255</td></td>	236 <td>AspProGlyLeuArgValAlaValGluValCysArgSerHisValArgIleAlaGluAla</td> <td>255</td>	AspProGlyLeuArgValAlaValGluValCysArgSerHisValArgIleAlaGluAla	255
Qy <td>902<td>---ACCATCAGTCTTGGCCCAAGGCCCTGCAGAACGCTACCTGCTGTGTCGCTATGAG</td><td>958</td></td>	902 <td>---ACCATCAGTCTTGGCCCAAGGCCCTGCAGAACGCTACCTGCTGTGTCGCTATGAG</td> <td>958</td>	---ACCATCAGTCTTGGCCCAAGGCCCTGCAGAACGCTACCTGCTGTGTCGCTATGAG	958
Db <td>256<td>AlaThrLeuLysPro---ProProPheLeuArgGlyArgTrpArgLeuValArgPheGlu</td><td>274</td></td>	256 <td>AlaThrLeuLysPro---ProProPheLeuArgGlyArgTrpArgLeuValArgPheGlu</td> <td>274</td>	AlaThrLeuLysPro---ProProPheLeuArgGlyArgTrpArgLeuValArgPheGlu	274
Qy <td>959<td>GACCTGCTCGAGCCCTGTGGCCAGACTTCCGGAATGTATGAATTCGTGGGATGGAA</td><td>1018</td></td>	959 <td>GACCTGCTCGAGCCCTGTGGCCAGACTTCCGGAATGTATGAATTCGTGGGATGGAA</td> <td>1018</td>	GACCTGCTCGAGCCCTGTGGCCAGACTTCCGGAATGTATGAATTCGTGGGATGGAA	1018
Db <td>275<td>AspLeuAlaArgGluProProLeuAlaGluIleArgAlaLeuTrpAlaPheThrGlyLeuSer</td><td>294</td></td>	275 <td>AspLeuAlaArgGluProProLeuAlaGluIleArgAlaLeuTrpAlaPheThrGlyLeuSer</td> <td>294</td>	AspLeuAlaArgGluProProLeuAlaGluIleArgAlaLeuTrpAlaPheThrGlyLeuSer	294
Qy <td>1019<td>TTCTTGGCCCATCTTTCAGACTGGGTGCATAACATCACCCGAGCAAGGCATGGGT---</td><td>1075</td></td>	1019 <td>TTCTTGGCCCATCTTTCAGACTGGGTGCATAACATCACCCGAGCAAGGCATGGGT---</td> <td>1075</td>	TTCTTGGCCCATCTTTCAGACTGGGTGCATAACATCACCCGAGCAAGGCATGGGT---	1075
Db <td>295<td>LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla</td><td>314</td></td>	295 <td>LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla</td> <td>314</td>	LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla	314
Qy <td>1076<td>---GACCACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTTCTCCAGGCTTGGCGC</td><td>1132</td></td>	1076 <td>---GACCACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTTCTCCAGGCTTGGCGC</td> <td>1132</td>	---GACCACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTTCTCCAGGCTTGGCGC	1132

Db 315 ArgArgGluAlaPheLysThrSerSerArgAsnAlaLeuAsnValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTGGCCATGAAAGGTTTTCGACTTCAGAAAGCGCTGGCGATGCCATGAAT 1192
Db 335 HisAlaLeuProPheAlaLysIleArgValGlnGluLeuCyEalaGlyAlaLeuGln 354
QY 1193 TTGCTGGGCTACCGCAGCTCAGATCTGAACAAGAACAGAGAAACCTGCTGATCTT 1252
Db 355 LeuLeuGlyTyArgProValTySerGluaspGluGlnArgAsnLeuAlaLeuAspLeu 374
QY 1253 CTGCTACCTGAGCTGCTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTTTGCTGCC 1312
Db 375 Val-----LeuPro-----ArgGlyLeuAsnGlyPhe----- 383
QY 1313 ACCTGGTGTGACGCTCAGTCACTTCTCTGAATGCTTCTGAGCCT 1357
Db 384 ThrTrpAla-----SerSerThrAlaSerHisPro 393

RESULT 9
ID ABB81554
AC ABB81554; standard; protein; 395 AA.
XX
XX ABB81554;
XX
DT 05-SEP-2002 (first entry)
XX Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22.
XX
XX Homo sapiens.
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
XX (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
XX
XX N-PSDB; ABN89506.
XX
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX.
XX Claim 13; Fig 1A-D; 69pp; English.
XX
XX The present sequence represents human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratinoplasty or keratectomy
XX
XX Sequence 395 AA;

Alignment Scores:
Pred. No.: 1.34e-82 Length: 395
Score: 1028.50 Matches: 224
Percent Similarity: 65.06% Conservative: 46

Best Local Similarity: 53.98% Mismatches: 108
Query Match: 27.49% Indels: 37
DB: 5 Gaps: 10
US-09-645-078-1 (1-2043) x ABB81554 (1-395)
QY 125 ACAATCTACTGCTTAAATAAATGAAGCTCCTGCTGTTTCTGGTTTCCAGATGCCATC 184
Db 12 ThrAlaLeuLeuLeuAlaGlnThrPheLeuLeuLeuPheLeuValSerArgPro----- 29
QY 185 TTGGCTCTATTCTTCCACATGTACACCCACAAATCAGCTCCCTCTCTATGATGAAGCAG 244
Db 30 -----GlyProSerSerProAlaGlyGly 37
QY 245 CCGAGCGCATGACAGCTGCTGCTTCTTCTTCTTCTGCGCTGCTGCTTCTTCTTCTGGG 304
Db 38 GluAlaArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGly 57
QY 305 CAGCTTTTGGGAGCAGCAGCAGATTTTCTACCTGATGGAGCCGCTGGCACGTGG 364
Db 58 GlnLeuPheAsnGlnHisProAspValPheTyLeuMetGluProAlaTrpHisValTrp 77
QY 365 ATGACCTTCAAGCAGAGCAGCCGCTGGATGCTGCGACATGGCTGGGGGATCTGATAGG 424
Db 78 ThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuValArg 97
QY 425 GCGCTTCTTGTGCGACATGAGCGTCTTTGATGCTATCATGGAACCTGGTCCCGGAGA 484
Db 98 SerValPheLeuCyAspMetAspValPheAspAlaTyLeu---ProTrpArgArgAsn 116
QY 485 CAGTCCAGCCTTTTTCAGTGGGAGACAGCCGGGCGCTGTGTTCTGCACCTGCTGAC 544
Db 117 LeuSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCySerSerProAlaCySer 136
QY 545 ATCATCCCAAGATGAATCATCCCGGGCTCACTGACAGCTCCTGCTGCGATCAACAG 604
Db 137 AlaPheProArgGlyAlaIleSerSerGluAlaValCysLysProLeuCyAlaArgGln 156
QY 605 CCCTTTGAGTGGTGGAGAGCCCTGCGCTCCTACAGCCAGCTGCTGCTCAAGAGAGTG 664
Db 157 SerPheThrLeuAlaArgGluAlaCysArgSerTySerHisValValLeuLysGluVal 176
QY 665 CGCTTCTTCAAGCTGAGTCCCTCTACCGCTGCTGGAAGACCCCTCCCTCAACCTGCAT 724
Db 177 ArgPhePheAsnLeuGlnValLeuTyProLeuLeuSerAspProAlaLeuAsnLeuArg 196
QY 725 ATGCTGACCTGGTCCGGAGCCCGCGCGTGTCCGTTCCCGAGAACCGACAAAGGA 784
Db 197 IleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAlaLys 216
QY 785 GATCTCATGATTGACAGTCGCATTGTGATGGGGCAGCATGAGCAAAATCTCAAGAGGAG 844
Db 217 AlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---GluAla 235
QY 845 GACCAACCTACTATGTATGATGAGTCTATCTGCCAAAGCCAGCTGGAGATCTACAG--- 901
Db 236 AspProGlyLeuArgValValArgGluValCysArgSerHisValArgIleAlaGluAla 255
QY 902 ---ACCATCCAGTCCCTTGGCCAAAGCCCTGACAGGAACGCTACTCTGTCGGCTATGAG 958
Db 256 AlaThrLeuLysPro---ProProPheLeuArgGlyArgTyArgLeuValArgPheGlu 274
QY 959 GACCTGGCTCGAGCCCTGTCGCCAGACTTCCCGAATGTATGAATTCGTGGATTTGAA 1018
Db 275 AspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyAlaPheThrGlyLeuSer 294
QY 1019 TTCCTGCCCATCTTCAGACCTGGGTGATACATACATCCCGAGGCAAGGCGATGGT--- 1075
Db 295 LeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCAAGCTTTCACAAATGCCAGGATGCCCTTAATGTCTCCAGCTTGGGCGC 1132
Db 315 ArgArgGluAlaPheLysThrSerSerArgAsnAlaLeuAsnValSerGlnAlaTrpArg 334

QY 362 TGGATGACCTTCAAGCAGACCCGCTGGATGCTGCACATGGCTGTGGGATCTGATA 421
Db 77 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeu 96
QY 422 CGGGCGCTTCTTGTGACATGAGCGCTTGTGATGCTCATATGGAACCTGGTCCCGG 481
Db 97 ArgSerValPheLeuCysAspMetAspValPheAspAlaTyrLeu--ProTyrArgArg 115
QY 482 AGACAGTCCAGCTCTTTCAGTGGGAGAACACCCGGGCGCTGTGTGACCTCCCTGT 541
Db 116 AsnIleSerAspLeuPheGlnTrpAlaValSerA-GAlaLeuCysSerProProValCys 135
QY 542 GACATCATCCCAAGATGAATCATCCCGCGGCTCACTGCAGGCTCTGTGCAGTCAA 601
Db 136 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLysProLeuCysAlaThr 155
QY 602 CAGCCCTTTGAGTGTGAGAGAGCGCTCCCTCCTACAGCCAGCTGGTGTCTAAGAG 661
Db 156 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 175
QY 662 GTGCGCTTCTCAACCTGAGCCCTCTACCGCTGCTGAAAGACCCCTCCCTCAACCTG 721
Db 176 ValArgPhePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 195
QY 722 CATATGTCGACCTGTCGGGACCCCGGCGGTTCCTGCTCCGAGAACGCAAAAG 781
Db 196 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAla 215
QY 782 GGAGATCTCATGATGACAGTCCGATGTGAGGGGCGAGCAGACAAAACCTCAAGAG 841
Db 216 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal--Glu 234
QY 842 GAGGACCAACCTTACTATGTGTCAGGTCTCTGCCAAGCCAGCTGAGATCTACAG 901
Db 235 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 254
QY 902 ---ACCATCCAGTCTTGGCCCAAGCCCTGCGAGACGCTACTGCTGTGGCTATGAG 958
Db 255 AlaAlaLeuHisLysProProPheLeuGlnAspArgTyrArgLeuValArgTyrGlu 274
QY 959 GACCTGGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGATTCGTGGGATGGAA 1018
Db 275 AspLeuAlaArgAspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGly 294
QY 1019 TTCTTCCCATCTTACAGCTGGGTGTCATACATCACCCGAGGCAAGGCGATGGT--- 1075
Db 295 LeuThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCACGCTTCCACACAAATGTCAGGATGCCCTTAATGTCTCCAGGCTTGGCGC 1132
Db 315 ArgArgGluAlaPheLysThrThrSerArgAspAlaLeuSerValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTTCCTTATGAAAGGTTCTCGACTTCAGAAAGCCCTGGGCGATGCCATGAAT 1192
Db 335 HisThrLeuProPheAlaLysIleArgArgValGlnGluLeuCysGlyGlyAlaLeuGln 354
QY 1193 TTGCTGGCTACCGCCACGTCAGATCTGACACAGACAGAAACCTGTGTGGATCTT 1252
Db 355 LeuLeuGlyTyrArgSerValHisSerGluLeuGlnArgAspLeuSerLeuAspLeu 374
QY 1253 CTGTCTACCT---GGACTGTCTCCCTGAGCAATCCACTAAGAGGGTTGAGAAG 1301
Db 375 Leu-LeuProArgGlyMetAspSerPheLysTrpAlaSerSerThrGluLys 391

US-09-645-078-1 (1-2043) x AAU11275 (1-395)
QY 128 ATGCTACTGCTATAA-----AAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCC 181
Db 1 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuLeuMetValGlnThrGly 20
QY 182 ATCTTGGCTCTATTCTTCCACATGTACAGCCACAAACATCATCCCTGCTCTATGAGGCA 241
Db 21 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 36

Mouse; betal,3GnT; betal,3-N-acetylglucosaminyltransferase; MECA-79;
L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
allergic contact dermatitis; lymphoma; chronic pneumonia; antileuker;
delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
antimicrobial; antipsoriatic; antidiabetic; dermatological;
antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
Mus musculus.
WO200185177-A1.
15-NOV-2001.
10-MAY-2001; 2001WO-US015452.
11-MAY-2000; 2000US-00569320.
(BURN-) BURNHAM INST.
Fukuda M, Yeh J, Hiraoka N;
WPI; 2002-075226/10.
N-PSDB; AAS16948.
New enzyme, useful for modifying acceptor molecule, comprises an isolated
L-selectin sulfotransferase-2 that directs expression of L-selectin
ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
GlcNAc 6-sulfotransferase.
Claim 28; Fig 10; 98pp; English.
The present invention provides a method of modifying an acceptor molecule
by contacting the acceptor with an isolated betal,3-N-
acetylglucosaminyltransferase (betal,3GnT) or an active fragment, where
betal,3GnT directs expression of a MECA-79 antigen. The invention also
provides a method of treating or preventing an L-selectin-mediated
condition by reducing the expression or activity of a betal,3GnT that
directs expression of a MECA-79 antigen. This can be done by
administering to the subject an oligosaccharide L-selectin antagonist
that inhibits binding of L-selectin to a MECA-79 antigen, for example by
administering antibody material that specifically binds betal,3GnT,
and/or a betal,3GnT antisense nucleic acid molecule. L-selectin
sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
Alternatively, the expression or activity of LSST-2 or its active
fragment can be reduced in combination with reducing the expression or
activity of betal,3GnT. The method is useful for treating L-selectin
mediated conditions such as Crohn's disease and ulcerative colitis,
inflammatory disorders of the skin such as allergic contact dermatitis,
psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
hypersensitivity reactions, diabetes and hyperplastic thymus. This
sequence represents mouse I-GlcNAc6ST
Sequence 395 AA;
Alignment Scores:
Pred. No.: 7 67e-82 Length: 395
Score: 1020.00 Matches: 221
Percent Similarity: 66.58% Conservatives: 44
Best Local Similarity: 55.53% Mismatches: 120
Query Match: 27.26% Indels: 13
DB: Gaps: 8
US-09-645-078-1 (1-2043) x AAU11275 (1-395)
QY 128 ATGCTACTGCTATAA-----AAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCC 181
Db 1 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuLeuMetValGlnThrGly 20
QY 182 ATCTTGGCTCTATTCTTCCACATGTACAGCCACAAACATCATCCCTGCTCTATGAGGCA 241
Db 21 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 36

QY 242 CAGCCGAGCGATGACGCTGCTGCTCTTCTTCTGCGCTCTGCTCTTCTTCTG 301
Db |||||
QY 37 LeuGlyGluAArgValHisValLeuValLeuSerTrpArgSerGlySerPheVal 56
Db |||||
QY 302 GGGCAGCTTTTGGGACACCCAGATGTTTCTACCTGATGAGCCGCGCTGCACGTG 361
Db |||||
QY 57 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal 76
QY 362 TGGATGACCTTCAAGCAGACGCGCTGGATGCTGCATGGCTGTGCGGATCTGATA 421
Db |||||
QY 77 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeu 96
QY 422 CGGCGCTTCTTGTGACATGAGCGCTTTGATGCTTACATGGAACCTGTCGCCGG 481
Db |||||
QY 97 ArgSerValPheLeuCysAspMetAspValPheAspAlaTyrLeu--ProTrpArgArg 115
QY 482 AGACAGTCCAGCTTCTTCACTGGGAGACAGCGCGCTGCTGCTGACCTGCCCTGT 541
Db |||||
QY 116 AsnIleSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProValCys 135
QY 542 GACATCATCCCAAGATGAATCATCCCGGGCTCACTGAGGCTCTGTGCGATCAA 601
Db |||||
QY 136 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLysProLeuCysAlaThr 155
QY 602 CAGCCCTTGTAGGTGGAGAGCGCTGCGCTCTTACAGCCAGCTGCTCTCAAGAG 661
Db |||||
QY 156 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 175
QY 662 GTGCGCTTCTTCAACCTCAGTCCCTCTACCGCTGTGTAAGACCCCTCCCTCAACTG 721
Db |||||
QY 176 ValArgPheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 195
QY 722 CATATGCTGACCTGCTGGAGACCGCGCGCTGCTGCTGCTGCGAGACGACAAAG 781
Db |||||
QY 196 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluGlnThrAla 215
QY 782 GGAGATCATCATGATGACATGCGCTGCTGATGGGCGAGCATGACGAAAACTCAAGAAG 841
Db |||||
QY 216 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal--Glu 234
QY 842 GAGGACCAACCTTACTATGTATGAGTCACTCTCCAAAGCCAGCTGGAATCTACAG 901
Db |||||
QY 235 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 254
QY 902 ---ACATCCAGCTTGTGCCAAGCGCTGAGGAACTGCTGCTGCTGCTGCTGCTGAG 958
Db |||||
QY 255 AlaAlaLeuHisLysProProPheLeuGlnAspArgTyrArgLeuValArgTyrGlu 274
QY 959 GACCTGGCTGAGCCCTGCTGCGCCAGACTTCCGAAATGTATGAATTCGTGGGATGAA 1018
Db |||||
QY 275 AspLeuAlaArgAspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGly 294
QY 1019 TTCTTGGCCCATCTTCAGACCTGGGTGATACATCACCGGAGCAAGGCGATGGT--- 1075
Db |||||
QY 295 LeuThrProGlnLeuGlnTrpIleHisAsnIleThrHisGlySerGlyProGlyAla 314
QY 1076 ---GACCACGCTTCCACAAATCCAGGATGCTTAAATGCTCCAGCTTGGCGC 1132
Db |||||
QY 315 ArgArgGluAlaPheLysThrSerArgAspAlaLeuSerValSerGlnAlaTrpArg 334
QY 1133 TGGTCTTTCCTATGAAGAAGTTTCTCGACTTCAGAAAGCTGCGGATGCCATGAAT 1192
Db |||||
QY 335 HisThrLeuProPheAlaLysIleArgArgValGlnGluLeuCysGlyValAlaLeuGln 354
QY 1193 TTGCTGGGCTACCGCACCTCAGATCTGAACAGAACAGAACTGTGTGCTGATCTT 1252
Db |||||
QY 355 LeuLeuGlyTyrArgSerValHisSerGluLeuGlnArgAspLeuSerLeuAspLeu 374
QY 1253 CTGCTTACCT---GGACTGCTCCCTGAGCAATCCACTAGAGGGTTGAGAG 1301
Db |||||
QY 375 Leu-LeuProArgGlyMetAspSerPheLysTrpAlaSerSerThrGluLys 391
Db |||||

RESULT 13

ABB81555
ID ABB81555 standard; protein; 395 AA.
XX
AC ABB81555;
XX
XX 05-SEP-2002 (first entry)
XX
DE Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
XX Homo sapiens.
OS Mus musculus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 10 /label= Ala, Thr, Val
FT FT Misc-difference 13 /label= Ala, Val, Ser
FT FT Misc-difference 20 /label= Phe, Cys, Gly
FT FT Misc-difference 39 /label= Ala, Asp, Glu
FT FT Misc-difference 96 /label= Val, Met, Ile
FT FT Misc-difference 142 /label= Ala, Thr, Asn
FT FT Misc-difference 147 /label= Ala, Asp, Glu
FT FT Misc-difference 159 /label= Thr, Ser, Gly
FT FT Misc-difference 238 /label= Gly, His, Arg
FT FT Misc-difference 294 /label= Ser, Thr, Gly
FT FT Misc-difference 371 /label= Ala, Thr, Ser
FT FT Misc-difference 380 /label= Leu, Pro, Met
FT FT Misc-difference 382 /label= Gly, His, Ser
FT FT Misc-difference 384 /label= Thr, Ser, Lys
FT FT Misc-difference 390 /label= Ala, Glu
FT FT Misc-difference 391 /label= Ser, Lys
FT FT Misc-difference 392 /label= His, Gln
FT FT Misc-difference 394 /label= Arg, Glu
FT FT Misc-difference 395 /label= Asn, Ser
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
XX PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
XX PA (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
XX WPI; 2002-507643/54.
DR

DR WPI; 2001-138471/14.
 DR N-PSDB; AAD02697, AAD02698, AAD02699.
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 PS Claim 3; Fig 1; 128pp; English.
 CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
 CC alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
 CC membrane protein useful for inhibiting a binding event between a selectin
 CC and a selectin ligand, which comprises contacting the selectin with a non
 CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation. Note: The present sequence is also shown in
 CC sequence listing (page no: 56) but lacks four nucleotides at its 3' end
 XX Sequence 390 AA;
 SQ
 Alignment Scores:
 Pred. No.: 7,32e-81 Length: 390
 Score: 1009.00 Matches: 206
 Percent Similarity: 66.41% Conservative: 57
 Best Local Similarity: 52.02% Mismatches: 105
 Query Match: 26.96% Indels: 28
 DB: 4 Gaps: 7
 US-09-645-078-1 (1-2043) x AAY72639 (1-390)
 QY 122 AGCACAATCTACTGCTAAATAAAGCTCTGCTGTTTCTGGTTTCCAGATGCC 181
 Db 12 ThrValLeuLeuLeuAlaGlnThrCysLeuLeuLeuPheIleIleSerArgPro--- 30
 QY 182 ATCTGGCTTATCTTCCACATGATACAGCCAAACATAGCTCCCTCTCTATGAGGCA 241
 Db 31 -----GlyProSerSerProAlaGly 37
 QY 242 CAGCCGAGCGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 301
 Db 38 GlyGluAspArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheLeu 57
 QY 302 GGGCAGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAGCCCGCTGGCAGCTG 361
 Db 58 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal 77
 QY 362 TGATGACCTTACGAGACAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
 Db 78 TrpThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuMet 97
 QY 422 CGGCGCGCTTCTTGGCCACATGAGCGCTTTGATGCTTACCTGATGAGAACTGGTCCCGG 481
 Db 98 ArgSerIlePheLeuCysAspMetAspValPheAspAlaTyrMet---ProGlnSerArg 116
 QY 482 AGCAGCTCCAGCTCTTTTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
 Db 117 AsnLeuSerAlaPhePheAsnTrpAlaThrSerArgAlaLeuCysSerProProAlaCys 136
 QY 542 GACATCATCCCAAGAAAGTAAATCATCCCGCGCTCACTGCGAGGCTCTGTCGACCTGCT 601
 Db 137 SerAlaPheProArgGlyThrIleSerLysGlnAspValCysLysThrLeuCysThrArg 156
 QY 602 CAGCCCTTTCAGGTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661

Db 157 GlnProPheSerLeuAlaArgGluAlaCysArgSerTyrSerHisValValLeuLysGlu 176
 QY 662 GTGCGCTTCTTTCACACTGCAGTCCCTCTACCGCTGCTGTAAGACCCCTCCCTCAACCTG 721
 Db 177 ValArgPhePheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 196
 QY 722 CATATCGTGCACCTGCTCCGGAGCCCGCGCGCTGTTCCGTTCCCGAGAACGCACAAAG 781
 Db 197 ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluAlaAlaGly 216
 QY 782 GGAGATCTCATGATTGACAGTGCATTTGATGGGCGAGCATGAGCAAAACTCAAGAAG 841
 Db 217 ProIleLeuAlaArgAspAsnGlyIleValLeuGlyThrAsn---GlyLysTrpValGlu 235
 QY 842 GAGGACCAACCTTACTATGTGTGAGTGCAGTCTATCCCAAGCCAGCTGAGATCTTACAG 901
 Db 236 AlaAspProHisLeuArgLeuIleArgGluValCysArgSerHisValArgIleAlaGlu 255
 QY 902 -----ACCATCCAGTCTTGGCCAAAGCCCTGCAAGAACCGTACTCTCTGTGGCTAT 955
 Db 256 AlaAlaThrLeuLysPro---ProProPheLeuArgGlyArgTyrArgLeuValArgPhe 274
 QY 956 GAGGACCTGCTGAGCCCTGCTGCGCCAGACTTCCCGAATGTATCAATTCTGGGATG 1015
 Db 275 GluAspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThrGlyLeu 294
 QY 1016 GAATCTTGGCCATCTTCAGACCTGGGTGCTATACATCACCCGAGGCAAGGCGATGGT 1075
 Db 295 ThrLeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyIleGly 314
 QY 1076 GAC-----CAGCTTTCCACAAATGCCAGGATGCCCTTAATGTCTCCCGAGCTGG 1129
 Db 315 LysProIleGluAlaPheHisThrSerSerArgAsnAlaArgAsnValSerGlnAlaTrp 334
 QY 1130 CGCTGCTTTCGCTATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCGATGCCATG 1189
 Db 335 ArgHisAlaLeuProPheThrLysIleLeuArgValGlnGluValCysAlaGlyAlaLeu 354
 QY 1190 AATTTCTGGCTACCGCCAGCTCAGATCTGAACAGAACAGAGAACTGTGTGCTGAT 1249
 Db 355 GlnLeuLeuGlyTyrArgProValTyrSerAlaAspGlnArgAspLeuThrLeuAsp 374
 QY 1250 CTTCGTG-----TCTACTGGACTGTCCCTGAG 1276
 Db 375 LeuValLeuProArgGlyProAspHisPheSerTrpAlaSerProAsp 390
 RESULT 15
 ID ABB81556
 XX ABB81556 standard; protein; 390 AA.
 AC ABB81556;
 DT
 XX 05-SEP-2002 (first entry)
 DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
 XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; G1CNAC6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 XX Homo sapiens.
 OS
 XX US2002061562-A1.
 PN
 XX 23-MAY-2002.
 PD
 XX 09-AUG-2001; 2001US-00927602.
 PF
 XX 11-AUG-2000; 2000US-00638211.
 PR
 XX 11-AUG-2000; 2000US-0325773P.
 XX (FUKUJ) FUKUDA M N.
 PA (AKAW) AKAWA T O.

Search completed: May 6, 2004, 11:01:02
Job time : 113.5 secs

XX PI Fukuda MN, Akama TO;
 XX WPI; 2002-507643/54.
 XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 XX useful for treatment, monitoring and diagnosis of macular corneal
 XX dystrophy.

XX Example 5; Fig 2A-B; 69pp; English.

XX The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (II) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratinoplasty or keratectomy. The present sequence
 CC represents human intestinal N-acetylglucosamine-6- sulfotransferase,
 CC which is given in comparison with (I) in the exemplification of the
 CC present invention

XX Sequence 390 AA;

Alignment Scores:
 Pred. No.: 7, 32e-81 Length: 390
 Score: 1009.00 Matches: 206
 Percent Similarity: 66.41% Conservative: 57
 Best Local Similarity: 52.03% Mismatches: 105
 Query Match: 26.96% Indels: 28
 DB: Gaps: 7

US-09-645-078-1 (1-2043) x ABB81556 (1-390)

QY	122	AGCAACAATGCTACTCCTTAAATAATGAGCTCTGCTGTTCCTGGTTCCCGATGCCG	181
Db	12	ThrValLeuLeuAlaGlnThrThrCysLeuLeuPheIleIleSerArgPro---	30
QY	182	ATCTTGCGCTATTCTTCACATGTACAGCCACCAATCAGCTCCCTGTCTATGAAGCCA	241
Db	31	-----GlyProSerProAlaGly	37
QY	242	CAGCCGGAGCGGATGCAGTCTGGTCTGTCTTCCTGGCGCTCTGGCTCTTTTGTG	301
Db	38	GlyGluAspArgValHisValLeuValLeuSerTrpArgSerGlySerSerPheLeu	57
QY	302	GGCGAGCTTTTGGCAGCACCCAGATGTTTTCTACCTGATGAGCCCGCTGGCACGNG	361
Db	58	GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal	77
QY	362	TGGATGACCTTCAAGCAGACGACCCTCGGTGGATGCTGCACATGGCTGTGGGGATCTGATA	421
Db	78	TrpThrThrLeuSerGlnGlySerAlaAlaThrLeuHisMetAlaValArgAspLeuMet	97
QY	422	CGGGCGCTCTTTCTGGCAGATGAGCGCTCTTTTATGCCTACATGNACTGTGTCCCCGG	481
Db	98	ArgSerIlePheLeuCysAspMetAspValPheAspAlaTyrMet---ProGlnSerArg	116
QY	482	AGACAGCTCCAGCTCTTTTCAGTGGGAGAACAGCCGGCCCTGTGTCTTCGACCTGCCTGT	541
Db	117	AsnLeuSerAlaPhePheasnTrpAlaThrSerArgAlaLeuCysSerProProAlaCys	136
QY	542	GACATCATCCACAGATGAATCATCCCCCGGGCTCATCTGACGGCTCTGTGTGATCAA	601
Db	137	SerAlaPheProArgGlyThrIleSerLysGlnAspValCysLysThrLeuCysThrArg	156
QY	602	CAGCCCTTTGAGTGTGGAGAAGCGCTCCGCTCTCTACAGCACAGCTGGTGTCAAGGAG	661
Db	157	GlnProPheSerLeuAlaArgGluAlaCysArgSerTyrSerHisValLeuLeuLysGlu	176
QY	662	GTGCGCTTCTTCAACTGAGTCCCTCTACCCGCTGTGAAAGACCCCTCCCTCAACCTG	721

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 11:05:49 ; Search time 93.5 Seconds
(without alignments)
12129.816 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 3742
Sequence: 1 gaattccattgtgttgga.....tgggattctataaaaaaaa 2043

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Dlelop 6.0 , Dlext 7.0

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 2281346

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/p/US09645078/runat_06052004_104601_21872/app_query.fasta_1.2183
-DB=Published Applications_AA -QFM=fastan -SUFFIX=rapb -MINWATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALG=GN15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09645078 @CGN 1.1.13 @runat_06052004_104601_21872
-NCPD=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRA=10 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Result No.	Score	Match	Length	ID	Description
1	386	54.5	2038	9	US-09-816-825-2
2	386	54.5	2038	13	US-10-007-362-1
3	386	53.0	1984	14	US-10-427-631-11
4	418	27.5	1030.5	9	US-09-927-602-5
5	395	27.5	1028.5	12	US-09-927-602-2
6	395	27.2	1017.5	9	US-10-258-080-5
7	395	27.0	1009	9	US-09-927-602-4
8	171	21.9	821	9	US-09-927-602-8
9	483	16.0	598.5	14	US-10-212-933-2
10	484	15.7	587.5	14	US-10-212-933-4
11	531	15.7	587.5	12	US-09-833-790-255
12	531	14.7	549	12	US-10-211-462-97
13	411	14.7	549	14	US-10-021-660-128
14	169	13.9	519.5	9	US-09-927-602-6
15	169	13.7	511.5	9	US-09-927-602-7
16	481	13.5	504.5	12	US-10-087-192-123
17	488	13.0	488	12	US-10-087-192-126
18	479	9.0	336	12	US-09-927-602-9
19	169	9.0	335.5	9	US-09-927-602-10
20	179	7.8	293	9	US-09-927-602-11
21	174	7.8	293	9	US-09-927-602-12
22	19695	5.5	207.5	15	US-10-084-846A-3
23	19695	5.4	201	15	US-10-084-846A-3
24	19608	5.4	198.5	15	US-10-084-846A-8
25	19662	5.2	193	15	US-10-084-846A-6
26	187.5	5.0	19725	15	US-10-084-846A-4
27	182.5	4.9	19723	15	US-10-084-846A-5
28	178	4.8	178	12	US-10-425-114-56601
29	173	4.6	173	15	US-10-084-846A-8
30	172.5	4.7	172.5	15	US-10-084-846A-7
31	170.5	4.6	170.5	14	US-10-213-509-5
32	170.5	4.6	170.5	15	US-10-085-198-2
33	170.5	4.6	170.5	15	US-10-084-846A-5
34	161	4.4	161	15	US-10-084-846A-4
35	160.5	4.3	160.5	12	US-10-425-114-56601
36	156	4.2	156	15	US-10-084-846A-3
37	155	4.1	155	15	US-10-353-690-36
38	147	4.0	147	15	US-10-342-331-48
39	147	4.0	147	15	US-10-342-331-50
40	147	4.0	147	14	US-10-156-761-10907
41	147	4.0	147	15	US-10-342-331-49
42	146.5	3.9	146.5	14	US-10-184-644-249
43	146.5	3.9	146.5	14	US-10-184-634-249
44	145	3.9	145	15	US-10-260-937-16
45	145	3.9	145	12	US-09-918-715-261

ALIGNMENTS

RESULT 1
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ IDS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Alignment Scores: 1.49e-157 Length: 386
Pred. No.: 1.49e-157 Length: 386

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Match Length DB ID Description

Score: 2038.00 Matches: 386
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 54.46% Indels: 0
 DB: 9 Gaps: 0

US-09-645-078-1 (1-2043) x US-09-816-825-2 (1-386)

QY 128 ATGCTACTGCTAAATAAAGTAAAGCTCTGCTGTTCTGTTCTCCAGATGGCCATCTTG 187
 Db 1 MetLeuLeuProLysMetLysLeuLeuPheLeuValSerGlnMetAlaLeu 20
 QY 188 GCTCTATTCTCCACATGATACAGCCACACATCAGCTCCCTGCTGATGAGGACAGCC 247
 Db 21 AlaLeuPhePheHisMetTyrSerHisAgnlleSerSerLeuSerMetLysAlaGlnPro 40
 QY 248 GAGCGCATGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307
 Db 41 GluArgMetHisValLeuValLeuSerSerTyrPargSerGlySerPheValGlyGln 60
 QY 308 CTTTTTGGGAGCACCAGATGTTTCTACCTGATGAGCCGCCCTGGACCGCTGGATG 367
 Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTyrHisValTyrMet 80
 QY 368 ACCTTCAAGCAGACCGCTCGATGCTGCACATGCTGCTGCTGCTGCTGCTGCTGCTG 427
 Db 81 ThrPheLysGlnSerThrAlaTyrMetLeuHisMetAlaValArgAspLeuAla 100
 QY 428 GTCTTCTTGTGCGCATGAGCGTCTTTGTATGCTCCATGGAACCTGGTCCCGGAGACAG 487
 Db 101 ValPheLeuLysAspMetSerValPheAspAlaTyrMetGluProGlyProArgArgGln 120
 QY 488 TCCAGCCTTTTTCAGTGGGAGAACAGCCGGCCCTGCTGTTCTGCACCTGCTGTCATC 547
 Db 181 PhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsnLeuHisIle 200
 QY 728 GTGCACCTGTGCGGAGACCCCGCGCGCTGTTCCGTTCCCGAGAACGACAAAGGAGAT 787
 Db 201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAsp 220
 QY 788 CTCTATGATTGACAGTCGATTGTGATGGGCGAGCATGAGCAAAATCTCAAGAGGAGGAC 847
 Db 221 LeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAsp 240
 QY 848 CAACCTCTATGTGATGCGAGTTCATGTCGCAAGCGAGTGGAGATCTCAAGACCATC 907
 Db 241 GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
 QY 908 CAGTCTTGTCCCAAGGCGCTGACAGACGCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 967
 Db 261 GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla 280
 QY 968 CGAGCCCTGTGGCCAGACTTCCGGAATGATGAATTCGTGGATGGAAATCTTGTGCC 1027
 Db 281 ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro 300
 QY 1028 CATCTTCAGACCTGGGTCATACATCACCCGAGGCAAGGCGATGGTGACACCGCTTTC 1087
 Db 301 HisLeuGlnThrTyrPheValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
 QY 1088 CACAAATAAGCCAGGATGCCCTTAATGTCTCCAGGCTGGCGCTGCTTGTGCCCTAT 1147

Db 321 HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTyrArgTyrSerLeuProTyr 340
 QY 1148 GAAAGAGTTTCTCGACTTCAGAAAGCTGCGGATGCCATGAATTTGCTGGGCTACCGC 1207
 Db 341 GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrArg 360
 QY 1208 CACGTCAGATCTGAACAAGAACAGAGAAACCTGTTCTGCGATCTTCTGTACTCGACT 1267
 Db 361 HisValArgSerGluGlnGluArgAsnLeuLeuAspLeuLeuSerThrTyrThr 380
 QY 1268 GTCCCTGAGCAATCCAC 1285
 Db 381 ValProGluGlnIleHis 386
 RESULT 2
 US-10-007-262-1
 ; Sequence 1, Application US/10007262
 ; Publication No. US20020164748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/10/007,262
 ; PRIORITY FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 386
 ; TYPE: PRN
 ; ORGANISM: H. sapiens
 ; US-10-007-262-1

Alignment Scores:
 Pred. No.: 1,49e-157 Length: 386
 Score: 2038.00 Matches: 386
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 54.46% Indels: 0
 DB: 13 Gaps: 0

US-09-645-078-1 (1-2043) x US-10-007-262-1 (1-386)

QY 128 ATGCTACTGCTAAATAAAGTAAAGCTCTGCTGTTCTGTTCTCCAGATGGCCATCTTG 187
 Db 1 MetLeuLeuProLysLysMetLysLeuLeuPheLeuValSerGlnMetAlaLeu 20
 QY 188 GCTCTATTCTCCACATGATACAGCCACACATCAGCTCCCTGCTATGAAGGACAGCC 247
 Db 21 AlaLeuPhePheHisMetTyrSerHisAgnlleSerSerLeuSerMetLysAlaGlnPro 40
 QY 248 GAGCGCATGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307
 Db 41 GluArgMetHisValLeuValLeuSerSerTyrPargSerGlySerPheValGlyGln 60
 QY 308 CTTTTTGGGAGCACCAGATGTTTCTACCTGATGAGCCGCCCTGGACCGCTGGATG 367
 Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTyrHisValTyrMet 80
 QY 368 ACCTTCAAGCAGACCGCTCGATGCTGCACATGCTGCTGCTGCTGCTGCTGCTGCTG 427
 Db 81 ThrPheLysGlnSerThrAlaTyrMetLeuHisMetAlaValArgAspLeuAla 100
 QY 428 GTCTTCTTGTGCGCATGAGCGTCTTTGTATGCTCCATGGAACCTGGTCCCGGAGACAG 487
 Db 101 ValPheLeuLysAspMetSerValPheAspAlaTyrMetGluProGlyProArgArgGln 120
 QY 488 TCCAGCCTTTTTCAGTGGGAGAACAGCCGGCCCTGCTGTTCTGCACCTGCTGTCATC 547

```
Db 121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuCysSerAlaProAlaCysAspIle 140
Qy 548 ATCCCAAGATGAATCATCCCGGGCTCACTGCAGGCTCTGTGCTCAACAGCCC 607
Db 141 IleProGlnAspGluIleProArgAlaHisCysArgLeuLeuCysSerGlnGlnPro 160
Qy 608 TTGAGGTGTGTGAGAGGCTGCGCTCTACAGCCAGCGTGTCTCAAGAGGTCGC 667
Db 161 PheGluValValGluLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
Qy 668 TTCTTCAACCTGAGTCCCTTACCCGCTGCTGAAGACCCCTCCCTCAACCTGCATATC 727
Db 181 PhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsnLeuHisIle 200
Qy 728 GTGCACCTGTGTCCGGGACCCCGGCGGTGTCCGTTCCTCCGAGAACCCACAAAGGGAGAT 787
Db 201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAsp 220
Qy 788 CTCATGATTGACGTCCGATTGTGATGGGCGAGCATGAGCAAAACCTCAAGAGGAGAC 847
Db 221 LeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAsp 240
Qy 848 CAACCTCTACTATGTGATCAGGTCTATCTGCCAAAGCCAGCTGGAGATCTCAAGACCATC 907
Db 241 GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
Qy 908 CAGTCTTGGCCAGGCGCTGAGGAAGCTACCTGCTGTGTGCGCTATGAGGACCTGCT 967
Db 261 GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValAlaGlyrGluAspLeuAla 280
Qy 968 CGAGCCCTGTGGCCGACACTTCCCGAATGTATGAATTCGTGGGATTGGAAATCTTGCCC 1027
Db 281 ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro 300
Qy 1028 CATCTTCAGACCTGGGTGCATACATCACCCGAGGCGAGGCGATGGGTGACACGCTTTC 1087
Db 301 HisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
Qy 1088 CACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGCTTGGCGTGTCTTTCCTTAT 1147
Db 321 HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTyr 340
Qy 1148 GAAAGGTTTTCGACTTCAGAAAGCTGTGGCGATGCCATGAATTTGCTGGGTACCGC 1207
Db 341 GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLysGlyTyrArg 360
Qy 1208 CAGCTCAGATCTGAACAGACAGAGAACCTGTGCTGGATCTTCTCTACCTGGACT 1267
Db 361 HisValArgSerGluGlnGluGlnArgAsnLeuLeuLeuAspLeuLeuSerThrTrpThr 380
Qy 1268 GTCCCTGAGCAATCCAC 1285
Db 381 ValProGluGlnIleHis 386
```

```
RESULT 3
US-10-427-631-11
; Sequence 11, Application US/10427631
; Publication No. US20030175923A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: CORLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT APPLICATION NUMBER: US/10/427,631
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
```

```
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: m5sc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
; US-10-427-631-11

Alignment Scores:
Pred. No.: 386e-153 Length: 386
Score: 1984.00 Matches: 383
Percent Similarity: 99.48% Conservative: 2
Best Local Similarity: 98.97% Mismatches: 1
Query Match: 53.02% Indels: 2
DB: 14 Gaps: 0

US-09-645-078-1 (1-2043) x US-10-427-631-11 (1-386)
Qy 128 ATGCTACTCCCTAAATAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGCCCATTTG 187
Db 1 MetLeuLeuProLysLysMetLysLeuLeuPheLeuValSerGlnMetAlaIleLeu 20
Qy 188 GCTCTATTCTCCACATGTACAGCCACACATCAGCTCCCTCTCTATGAGGACAGCCC 247
Db 21 AlaLeuPhePheHisMetTyrSerHisAsnIleSerSerLeuSerMetLysAlaGlyPro 40
Qy 248 GAGCGATGACGCTGCTGCTGCTCTCTCTCTGCGCTCTGCTCTCTCTCTCTCTCT 307
Db 41 GluArgMetHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGlyGln 60
Qy 308 CTTTTTGGGACAGCCACCATGTTTCTACGTATGAGGAGCCCGCTGCGGATCTGATAC 367
Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrpMet 80
Qy 368 ACCTTCAAGCAGACAGCCCGCTGGATGTCACATGCTGCTGCGGATCTGATACGGGC 427
Db 81 ThrPheLysGlnSerThrAlaTrpMetLeuHisMetAlaValArgAspLeuIleArgAla 100
Qy 428 GTCTTCTTGTGCGACATGAGCGCTTTTGTAGTCTACATGGAACCTGCTCCCGGAGAC 487
Db 101 ValPheLeuCysAspMetSerValPheAspAlaTyrMetGluProGlyProArgArgGln 120
Qy 488 TCAGGCTCTTTTCAAGTGGGAGACAGCCGCGCTGTGTTCTGCACTGCTGTGACATC 547
Db 121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuCysSerAlaProAlaCysAspIle 140
Qy 548 ATCCCAAGATGAATCATCCCGGGCTCACTGCAGGCTCCCTGCGATCAACAGCCC 607
Db 141 IleProGlnAspGluSerSerProGlyLeuThrAlaGlySerCysAlaValAsnSerPr 160
Qy 608 TTGAGGTG-GTGAGAGAGCCCTGCGCTCTACAGCCAGCTGCTGTCTCAAGGAGGTGCG 666
Db 160 oLeuLysLeuLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
Qy 667 CTCTTCAACTGAGTCCCTCTACCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726
Db 180 gPhePheAsnLeuGlnSerLeuTyrProLeuLeuLysAspProSerLeuAsnLeuHisI 200
Qy 727 CGTGCACCTGCTCGGAGACCCCGGCGCTGTTCCGTTCCCGAGAGACGACAAAGGAGAG 786
Db 200 eValHisLeuValArgAspProArgAlaValPheArgSerArgGlnArgThrLysGlyAs 220
Qy 787 TCTCATGATTGACAGTCCGCTTGTGATGGGCGAGCATGAGCAAAATCTCAAGAGGAGGA 846
```

Db 220 pLeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAs 240
QY 847 CCAACCTACTATGTGATCAGTCTATCTGCCAAAGCCAGCTGGAGATCTCAAGACCAT 906
Db 240 pGlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrI 260
QY 907 CCAGTCCCTGCCCAGGCCCTCAGGAAGCTACCTGCTGTGCGCTATGAGGACCTGGC 966
Db 260 eGlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAl 280
QY 967 TCAGAGCCCTGTGGCCCAAGACTTCCCGAATGTATGAATTCGTGGGATTGGAATTCCTGCC 1026
Db 280 aarghlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPr 300
QY 1027 CCATCTTCAGACTGGGTGCATAAATCACCAGGAGGAGGCGATGGGTGACCGCTTT 1086
Db 300 ohisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaph 320
QY 1087 CCACACAAATGCCAGGATGCCCTTAATGTCTCCAGAGCTGGCGCTGTCTTCCCTCA 1146
Db 320 ehisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpA-gTrpSerLeuProTy 340
QY 1147 TGAAGAGTTTCTGACTCTCAGAAAGCTGTGGCATGCCATGCAATTTGCTGGGCTACG 1206
Db 340 rGluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrAr 360
QY 1207 CCAGTTCAGATCTGAACAAAGAACAGAGAAACCTGTGCTGGATCTCTCTACCTGGAC 1266
Db 360 gHisValArgSerGluGlnGluGlnArgAsnLeuLeuAspLeuLeuSerThrTrpTh 380
QY 1267 TGTCCCTCAGCAATCCAC 1285
Db 380 rValProGluGlnIleHis 386

RESULT 4

US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

Alignment Scores:
Pred. No.: 2,99e-75 Length: 418
Score: 1030.50 Matches: 221
Percent Similarity: 66.75% Conservative: 44
Best Local Similarity: 55.67% Mismatches: 120
Query Match: 27.54% Indels: 12
DB: 9 Gaps: 7

US-09-645-078-1 (1-2043) x US-09-927-602-5 (1-418)

QY 128 ATGCTACTGCTAAA-----AAATGAAGCTCTGCTGTCTGTTTCCAGATGGCC 181
Db 25 MetArgLeuProArgPheSerSerThrValMetLeuSerLeuLeuMetValGlnThrGly 44
QY 182 ATCTTGGCTCTATTCTTCCACATCATCAGCCACAAATCAGCTCCCTGTCTATGAAGGCA 241
Db 45 IleLeuValPhe-----LeuValSerArgGlnValProSer---SerProAlaGly 60

QY 242 CACCCGAGCGATGACAGTGTGTTCTGTCTTCTCTGGCGCTCTGGCTCTTCTTTTGG 301
Db 61 LeuGlyGluArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheVal 80
QY 302 GGGCAGCTTTTGGGAGCAGCAGCAGATGTTTCTACCTGATGAGCCCGCTGGCAGCG 361
Db 81 GlyGlnLeuPheSerGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisVal 100
QY 362 TGAATGACCTTCAAGCAGAGCAGCAGCGCTGATGCTGCACATGGCTGTGGGATCTGATA 421
Db 101 TrpAspThrLeuSerGlnGlySerAlaProAlaLeuHisMetAlaValArgAspLeuIle 120
QY 422 CGGGCGCTCTTCTGTGCGACATGAGCTCTTGTGCTTACATGGAACCTGCTCCCGG 481
Db 121 ArgSerValPheLeuLeuCysAspMetAspValPheAspAlaTyrLeu---ProTrpArg 139
QY 482 AGACAGTCCAGCTCTTTCAGTGGGAGAACAGCCGGCGCTGTGTCTGCACCTGCCCTG 541
Db 140 AsnIleSerAspLeuPheGlnTrpAlaValSerArgAlaLeuCysSerProValCys 159
QY 542 GACATCATCCCAAGATGAATCATCCCGCGCTCACTGAGGCTCTGTGCATCAA 601
Db 160 GluAlaPheAlaArgGlyAsnIleSerSerGluGluValCysLysProLeuCysAlaThr 179
QY 602 CAGCCCTTTGAGTGTGGAGAGGCTGCGCTCTCTACAGCCAGCTGTGTCTCAAGGAG 661
Db 180 ArgProPheGlyLeuAlaGlnGluAlaCysSerSerTyrSerHisValValLeuLysGlu 199
QY 662 GTGGCTCTTCTAACCTGACGTCCCTCTACCCGCTCTGAAAGACCCCTCCCTCAACCTG 721
Db 200 ValArgPheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeu 219
QY 722 CATATCGTCACCTGTGCGGACCCCGCGCTGTTCGTTCCCGAGAACGCACAAAG 781
Db 220 ArgIleValHisLeuValArgAspProAlaValLeuArgSerArgGluGlnThrAla 239
QY 782 GGAGATCTCATGATTGACAGTCCATGTGATGGGCGAGCATGAGCAAAACATCAAGAAG 841
Db 240 LysAlaLeuAlaArgAspAsnGlyIleValLeuGlyThrAsnGlyThrTrpVal---Glu 258
QY 842 GAGGACCAACCTACTATGTGATGATGAGTCCAGGTCATCTGCCAAAGCCAGCTGGAGATCTCAAG 901
Db 259 AlaAspProArgLeuArgValValAsnGluValCysArgSerHisValArgIleAlaGlu 278
QY 902 ACCATCAGCTCTTGGCCCAAGGCCCTGCAGAAAGCTACTGCTGTGCGCTATGAGGAC 961
Db 279 AlaLeuHisLysProProPheLeuGlnAspArgTyrArgLeuValArgTyrGluAsp 298
QY 962 CTGGCTCGAGCCCTGTGGCCAGACTTCCCGAATGTATGAATTCGTGGGATTGGAATTC 1021
Db 299 LeuAlaArgAspProLeuThrValIleArgGluLeuTyrAlaPheThrGlyLeuGlyLeu 318
QY 1022 TTGCCCCATCTTTCAGACTGGTGCATACATCACCGAGGCAAGGCATGGCT----- 1075
Db 319 ThrProGlnLeuGlnThrTrpIleHisAsnIleThrHisGlySerGlyProGlyAlaArg 338
QY 1076 GACCAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGG 1135
Db 339 ArgGluAlaPheLysThrThrSerArgAspAlaLeuSerValSerGlnAlaTrpArgHis 358
QY 1136 TCTTTGGCTATGAAAGTTCTTCGACTTCAGAAACCTGTGGCGATGCATGAATTCG 1195
Db 359 ThrLeuProPheAlaLysIleArgArgValGlnGluLeuCysGlyGlyAlaLeuGlnLeu 378
QY 1196 CTGGGCTACCCCGCAGCTCAGATCTGAACAAAGAACAGAGAAACCTGTGTGCTGATCTTCG 1255
Db 379 LeuGlyTyrArgSerValHisSerGluLeuGlnArgAspLeuSerLeuAspLeuLeu 398
QY 1256 TCTACCT---GGACTGTCCCTGAGCAAAATCCACTAAGAGGTTCCAGAG 1301
Db 399 -LeuProArgGlyMetAspSerPheLysTrpAlaSerSerThrGluLys 414

[illegible]

Accession	Species	Gene	Protein	Length	Score	Ident	Positives	Negatives	Indels	Gaps
197	Human	IL13	IL13	113	1028.50	65.06%	108	37	10	12
785	Human	GATC	GATC	113	1028.50	65.06%	108	37	10	12
217	Human	Ala	Ala	113	1028.50	65.06%	108	37	10	12
845	Human	GACCA	GACCA	113	1028.50	65.06%	108	37	10	12
236	Human	Asp	Asp	113	1028.50	65.06%	108	37	10	12
902	Human	ACC	ACC	113	1028.50	65.06%	108	37	10	12
256	Human	Ala	Ala	113	1028.50	65.06%	108	37	10	12
959	Human	GAC	GAC	113	1028.50	65.06%	108	37	10	12
275	Human	Asp	Asp	113	1028.50	65.06%	108	37	10	12
1019	Human	TTC	TTC	113	1028.50	65.06%	108	37	10	12
295	Human	Leu	Leu	113	1028.50	65.06%	108	37	10	12
1076	Human	GAC	GAC	113	1028.50	65.06%	108	37	10	12
315	Human	Arg	Arg	113	1028.50	65.06%	108	37	10	12
1133	Human	TGG	TGG	113	1028.50	65.06%	108	37	10	12
335	Human	His	His	113	1028.50	65.06%	108	37	10	12
1193	Human	TTG	TTG	113	1028.50	65.06%	108	37	10	12
355	Human	Leu	Leu	113	1028.50	65.06%	108	37	10	12
1253	Human	CTG	CTG	113	1028.50	65.06%	108	37	10	12
375	Human	Val	Val	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113	1028.50	65.06%	108	37	10	12
1313	Human	ACT	ACT	113	1028.50	65.06%	108	37	10	12
384	Human	Thr	Thr	113						

QY	542	GACATCATCCACAAAGATGAAATCATCCCGGGGCTCACTCAGGCTCCTCTGTGAGTCAA	601
Db	137	SerAlaPheProArgGlyThrIleSerIysGlnAspValCysLysThrLeuCysThrArg	156
QY	602	CAGCCCTTTGAGGTGGGAAAGCGCTGCGCTCCTACAGCCAGCTGGTGCTCAAGGAG	661
Db	157	GlnProPheSerLeuAlaArgGluAlaCysArgSerTyrSerHisValValLeuIysGlu	176
QY	662	GTGCGCTTCTCAACCTGCAGCTCCTCTACCCGCTGCTGAAGACCCCTCCTCAACCTG	721
Db	177	ValArgPhePheAsnLeuGlnValIleTyrProLeuLeuSerAspProAlaLeuAsnLeu	196
QY	722	CATATCGTGACCTGGTCCGGAGCCCGGGCCGTGCTCCGTCGAGACGCGACAAAG	781
Db	197	ArgIleValHisLeuValArgAspProArgAlaValLeuArgSerArgGluAlaGly	216
QY	782	GGAGATCTCATGATTGACAGTCGCAATGTGTGATGGCGCAGCATGAGCAAAACTCAAGAG	841
Db	217	ProIleLeuAlaArgAspAsnGlyIleValLeuGlyThrAsn--GlyLysTrpValGlu	235
QY	842	GAGGACCAACCTACTACTGATCGAGGTGATCTGCCNAAGCCAGCTGGAGATCTACAAG	901
Db	236	AlaAspProHisLeuArgLeuAlaArgGluValCysArgSerHisValArgIleAlaGlu	255
QY	902	-----ACCATCCAGTCTCTCCCAAGGCCCTGCAGGAACCTACCTGCTTGTCGCTAT	955
Db	256	AlaAlaThrLeuLysPro--ProProPheLeuArgGlyArgTyrArgLeuValArgPhe	274
QY	956	GAGGACCTGGTCTGAGCCCTGTGTGGCCAGACTTCCCGAATGTATGATTCGTGGGATG	1015
Db	275	GluAspLeuAlaArgGluProLeuAlaGluIleArgAlaLeuTyrAlaPheThr-GlyLeu	294
QY	1016	GAATTCTTGGCCCATCTTCAGACTGGTGTGATCAATCATCCAGCCAGCGCATGGT	1075
Db	295	ThrLeuThrProGlnLeuGluAlaTrpIleHisAsnIleThrHisGlySerGlyIleGly	314
QY	1076	GAC-----CACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCACGCTGG	1129
Db	315	LysProIleGluAlaPheHisThrSerSerArgAsnAlaArgAsnValSerGlnAlaTrp	334
QY	1130	CGTGTGCTTTGCCCTATGAAAAGGTTTCTCGATTCGAAAGCTGTGGCGATGCCCAT	1189
Db	335	ArgHisAlaLeuProPheThrLysIleLeuArgValGlnGluValCysAlaGlyAlaLeu	354
QY	1190	AATTGTGTGGGCTACCGCCAGTCAGATCTCAACAAAGACAGAAACCTCTTCGTGAT	1249
Db	355	GlnLeuLeuGlyTyrArgProValTyrSerAlaAspGlnGlnArgAspLeuThrLeuAsp	374
QY	1250	CTTCTG-----TCTACCTGGAGTGTCCCTGAG	1276
Db	375	LeuValLeuProArgGlyProAspHisPheSerTrpAlaSerProAsp	390

```

RESULT 9
US-09-927-602-8
; Sequence 8, Application US/09927602
; Patent No. US2002061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo Sapien

```

US-09-927-602-8

Alignment Scores:
 Pred. No.: 2,91e-58 Length: 171
 Score: 821.00 Matches: 171
 Percent Similarity: 60.00% Mismatches: 0
 Best Local Similarity: 60.00% Conservatives: 0
 Query Match: 21.94% Indels: 114
 DB: 9 Gaps: 2

US-09-645-078-1 (1-2043) x US-09-927-602-8 (1-171)

Qy	221	AGTCTCCGTCTATGAAGGACACGCCGAGCGATGCACGTGCTGTGTTCTGTCTCTCTCGG	280
Db	1	SerSerLeuSerMetLysAlaGlnProGluArgMetHisValLeuValLeuSerSerTyr	20
Qy	281	CGCTCTCGCTCTCTTTTGTGGGGCAGCTTTTGGCCAGCACCCAGATGTTTCTACTCG	340
Db	21	ArgSerGlySerSerPheValGlyGlnLeuPheGlyGlnHisProAspValPheTyrLeu	40
Qy	341	ATGAGAGCCGCCCTGGCAGCGTGTGGATGACCTTCAAGCAGACGACCGCCTGGATGCTGCAC	400
Db	41	MetGluProAlaTyrHisValTyrMetThrPheLys	52
Qy	401	ATGGCTGTGGGATCTGATACGGGCCGTCTTCTGTGGCAGATGAGCGTCTTTGATGCC	460
Db	52	-----	52
Qy	461	TACATGAACCTGTGTCGCCGGAGACAGTCCAGCGCTCTTTCAGTGGGAGAAACGCCGGGCC	520
Db	52	-----	52
Qy	521	CTGTGTTCTGCACCTGCCTGTGACATCATCCCAAGATGAAATCATCCCCCGGGCTCAC	580
Db	52	-----	52
Qy	581	TGCAGGCTCCTGTGTCAGTCAACAGCCCTTTGAGTGTGTGGAGAGGCGCTCCCGCTCTAC	640
Db	53	-----LysAlaCysArgSerTyr	58
Qy	641	AGCCAGCTGTGCTCAAGGAGGTGGCGTCTTCAACCTGCAGTCCCTCTACCGCTCGT	700
Db	59	SerHisValValLeuLysGluValArgPheAsnLeuGlnSerLeuTyrProLeuLeu	78
Qy	701	AAAGACCCCTCCCTCAACCTGCATATCGTGCACCTGTCTCGGACCCCGCGCGGTTC	760
Db	79	LysAspProSerLeuAsnLeuHisIleValHisLeuValArgAspProArgAlaValPhe	98
Qy	761	CGTTCCCGAGAACGCACAAAGGAGATCTCATGATTGACAGTGCATGTGTGGGGCAG	820
Db	99	ArgSerArgGluArgThrLysGlyAspLeuMetIleAspSer	112
Qy	821	CATGAGCAAAACTCAAGAAAGGAGGACCAACCTACTATGTGTGTCAGGTCTCTGCCAA	880
Db	112	-----	112
Qy	881	AGCCAGCTGGAGATCTACAAGACCATCCAGTCTCTTGCCCAAGGCCCTGCAGGAACGCTAC	940
Db	113	-----LysThrIleGlnSerLeuProLysAlaLeuGlnGluArgTyr	126
Qy	941	CTGCTTGTGGCTATGAGGACCTGTGCTCGAGCCCTGTGSCCCAGACTTCCCGAATGTAT	1000
Db	127	LeuLeuValArgTyrGluAspLeuAlaArgAlaProValAlaGlnThrSerArgMetTyr	146
Qy	1001	GAATTCTGTGGATTGGAATCTTGTGCCCATCTTTCAGACCTGGGTGGCATACATCACCCGA	1060
Db	147	GluPheValGlyLeuGluPheLeuProHisLeuGlnThrTyrValHisAsnIleThrArg	166
Qy	1061	GGCAAGGGCATGGGT	1075
Db	167	GlyLysGlyMetGly	171

RESULT 10


```

Db      263 -----LysProTyrAsnLeuAspValThrGlnLeuThrThrValCysGluAsp 278
QY      884 CAGCTGGAGATCTACAAGACCATCCAGTCTCTGCGCCCAAGCCCTGCAGGAACGCTACCTG 943
Db      279 PheSerAsnSerValSerThrGlyLeuMetArgProProtrpLeuLysGlyLysTyrMet 298
QY      944 CTTGTGGCTATAGGACCTGGCTGAGCCCTGAGCCCTGAGCCCTGAGCCCTGAGCCCTGAG 1003
Db      299 LeuValArgTyrGluAspLeuAlaArgAsnProMetLysThrGluGluIleTyrGly 318
QY      1004 TTTCTGGGATGGAAATCTTCTGCGCCCATCTTCAGACCTGGGTGCATACATCACCCGAGGC 1063
Db      319 PheLeuGlyIleProLeuAspSerHisValAlaArgTyrPheGlnAsnAsnThrArgGly 338
QY      1064 ---AAGGCATGGTGACCGCTTTCCACAAATGCCAGGATGCCCTTAATGTCTCC 1120
Db      339 AspProThrLeuGlyLysHisLysTyrGlyThr---ValArgAsnSerAlaAlaThrAla 357
QY      1121 CAGCTTGGCTGGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1180
Db      358 GluLysTrpArgPheArgLeuSerTyrAspIleValAlaPheAlaGlnAsnAlaCysGln 377
QY      1181 GATGCCATGAATTTGCTGGCTACCGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1240
Db      378 GlnValLeuAlaGlnLeuGlyTyrLysIleAlaAlaSerGluGluLeuLysAsnPro 397

RESULT 14
US-10-021-660-128
; Sequence 128, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1 Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-00071005
; CURRENT APPLICATION NUMBER: US/10/021-660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-128

Alignment Scores:
Pred. No.: 6,35e-36 Length: 411
Score: 549.00 Matches: 139
Percent Similarity: 50.12% Conservative: 64
Best Local Similarity: 34.32% Mismatches: 160
Query Match: 14.67% Indels: 42
DB: 14 Gaps: 10

US-09-645-078-1 (1-2043) x US-10-021-660-128 (1-411)
QY      143 AAAATGAAGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 193
Db      6 LysAlaValLeuLeuLeuAlaLeuAlaSerIleAlaIleGlnTyrThrAlaIleArgThr 25
QY      194 TTTCTCCATGATACAGGCACAAATCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253

```

```

Db      26 PheThrAlaLysSerPheHisThrCysProGlyLeuAlaGluAlaGlyLeuAlaGluArg 45
QY      254 ATG-----CACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 268
Db      46 LeuCysGluGluSerProThrPheAlaTyrAsnLeuSerArgLysThrHisIleLeuIle 65
QY      269 CTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 328
Db      66 LeuAlaThrThrArgSerGlySerPheValGlyGlnLeuPheAsnGlnHisLeuAsp 85
QY      329 GTTTTCTACCTGATGGAGCCGCTGCGACCTGGTGGATGACC-----TTCAAG 376
Db      86 ValPheTyrLeuPheGluProLeuTyrHisValGlnAsnThrLeuIleProArgPheThr 105
QY      377 CAGACACCGGC-----TGGATGCTGCACATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 424
Db      106 GlnGlyLysSerProAlaAspArgValMetLeuGlyAlaSerArgAspLeuLeuArg 125
QY      425 GCGCTCTCTTGTGGACATGAGCGTCTTTGATGCTACATGGAACCTGCTGCTGCTGCTGCTGCT 484
Db      126 SerLeuTyrAspCysAspLeuTyrPheLeuGlnAsnTyrIleLysProProValAsn 145
QY      485 CAGTCCAGC-----CTTTTTCAGTGGAGAACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
Db      146 HisThrThrAspArgIlePheArgGlyAlaSerArgValLeuCysSerArgProVal 165
QY      539 TGTGACATCATC---CCACAAGATCAATCATCCCGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCT 595
Db      166 CysAspProProGlyProAlaAspLeuValLeuGluGlyAspCysValArgLysCys 185
QY      596 ACTCAACAGCCCTTTGAGTGGTGGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 655
Db      186 GlyLeuLeuAsnLeuThrValAlaAlaGluAlaCysArgGluArgSerHisValAlaIle 205
QY      656 AAGGAGTGGCTTCTTCAACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 715
Db      206 LysThrValArgValProGluValAsnAspLeuArgAlaLeuValGluAspProArgLeu 225
QY      716 AACCTGCATATCGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
Db      226 AsnLeuLysValIleGlnLeuValArgAspProArgGlyIleLeuAlaSerArgSerGlu 245
QY      776 ACAAGGAGATCTCATGATGACAGTGCATGATGATGATGATGATGATGATGATGATGATGATG 835
Db      246 ThrPheArgAspThrTyrArgLeuTrpArgLeuTyrGlyThrGlyArg-----262
QY      836 AAGAAGGAGGACCAACCTACTATGTG-----ATGCAGTCTCATGCTGCTGCTGCTGCTGCTG 883
Db      263 -----LysProTyrAsnLeuAspValThrGlnLeuThrThrValCysGluAsp 278
QY      884 CAGCTGGAGATCTACAAGACCATCCAGTCTTGTGCGCCCAAGCCCTGCAGGAACGCTACCTG 943
Db      279 PheSerAsnSerValSerThrGlyLeuMetArgProProtrpLeuLysGlyLysTyrMet 298
QY      944 CTTGTGGCTATGAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 1003
Db      299 LeuValArgTyrGluAspLeuAlaArgAsnProMetLysThrGluGluIleTyrGly 318
QY      1004 TTTCTGGGATGGAAATCTTCTGCGCCCATCTTCAGACCTGGGTGCATACATCACCCGAGGC 1063
Db      319 PheLeuGlyIleProLeuAspSerHisValAlaArgTyrPheGlnAsnAsnThrArgGly 338
QY      1064 ---AAGGCATGGTGACCGCTTTCCACAAATGCCAGGATGCCCTTAATGTCTCC 1120
Db      339 AspProThrLeuGlyLysHisLysTyrGlyThr---ValArgAsnSerAlaAlaThrAla 357
QY      1121 CAGCTTGGCTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1180
Db      358 GluLysTrpArgPheArgLeuSerTyrAspIleValAlaPheAlaGlnAsnAlaCysGln 377
QY      1181 GATGCCATGAATTTGCTGGCTACCGCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1240
Db      378 GlnValLeuAlaGlnLeuGlyTyrLysIleAlaAlaSerGluGluLeuLysAsnPro 397

```

```
QY 1241 TTGCTGGATCTTCTG 1255
Db 398 SerValSerLeuVal 402
RESULT 15
US-09-927-602-6
; Sequence 6, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-6
Alignment Scores:
Pred. No.: 1,238-33 Length: 169
Score: 519.50 Matches: 113
Percent Similarity: 46.55% Conservative: 15
Best Local Similarity: 41.09% Mismatches: 34
Query Match: 13.88% Indels: 113
DB: 9 Gaps: 3
US-09-645-078-1 (1-2043) x US-09-927-602-6 (1-169)
QY 251 CGATGACAGTGGTCTGCTCTCTGCGGCTCTGCTCTCTTTTGTGGGAGAGCTT 310
Db 8 ArgValHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGlyLeu 27
QY 311 TTGGGCGACACCCAGATGTTTCTACCTGATGAGCGCGCTGGCAGCTGTGGATGAC 370
Db 28 PheAsnGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrpThr 47
QY 371 TTCAGCAGAGACCGCTGGATGTCACATGGCTGTGGGGATCTGATACGGGCGGTC 430
Db 47 ----- 47
QY 431 TTCTTGTGCGACATGAGGCTCTTTGATGCTACATGGAACCTGGTCCCGGAGACAGTCC 490
Db 47 ----- 47
QY 491 AGCCTCTTTCAGTGGGAGAACAGCCGGGCGCTGTCTTCTGCACTGCTGTGACATCATC 550
Db 47 ----- 47
QY 551 CCACAAGATGAATCATCCCCGGGCTCACTGCGAGGCTCTGTGCGAGTCAACAGCCCTTT 610
Db 47 ----- 47
QY 611 GAGTGGTGGAGAGCGCTGCGCTCTACAGCCACGCTGCTCAAGAGGTGCGCTTC 670
Db 48 -----LeuSerGluAlaCysArgSerTyrSerHisValValLeuLysGluValArgPhe 65
QY 671 TTCACCTGTCAGTCTCTACCGCTGCTGAAAGACCGCTCCCTCACTGATATCGTG 730
Db 66 PheAsnLeuGlnValLeuTyrProLeuLeuSerAspProAlaLeuAsnLeuArgIleVal 85
QY 731 CACCTGGTCCGGGACCCCGGCGCTGTTCGTTCCCGAGAACGACAAAGGAGATCTC 790
Db 86 HisLeuValArgAspProAlaValLeuArgSerArgGluGlnThrAlaLysAlaLeu 105
```

Search completed: May 6, 2004, 11:23:04
Job time : 109.5 secs

C 13	165.5	4.5	265	4	US-09-252-991A-32896	Sequence 32896, A
C 14	162.5	4.3	335	4	US-09-252-991A-23948	Sequence 23948, A
C 15	161.5	4.4	1034	4	US-09-252-991A-28921	Sequence 28921, A
C 16	160	4.3	232	4	US-09-252-991A-19777	Sequence 19777, A
C 17	159.5	4.3	421	4	US-09-252-991A-30742	Sequence 30742, A
C 18	157.5	4.3	504	4	US-09-252-991A-28242	Sequence 28242, A
C 19	157	4.2	341	4	US-09-252-991A-32424	Sequence 32424, A
C 20	155.5	4.2	465	4	US-09-252-991A-20576	Sequence 20576, A
C 21	155.5	4.2	301	4	US-09-252-991A-30700	Sequence 30700, A
C 22	155.5	4.2	492	4	US-08-468-995-12	Sequence 12, Appl
C 23	154.5	4.2	315	4	US-09-252-991A-20553	Sequence 20553, A
C 24	154	4.2	556	4	US-09-252-991A-22670	Sequence 22670, A
C 25	154	4.2	1089	4	US-09-252-991A-20334	Sequence 20334, A
C 26	153.5	4.1	369	4	US-09-252-991A-25394	Sequence 25394, A
C 27	153.5	4.1	448	4	US-09-252-991A-24066	Sequence 24066, A
C 28	152.5	4.1	215	4	US-09-252-991A-28157	Sequence 28157, A
C 29	152	4.1	582	4	US-09-252-991A-26182	Sequence 26182, A
C 30	151.5	4.1	170	4	US-09-252-991A-17086	Sequence 17086, A
C 31	151.5	4.1	250	4	US-09-252-991A-16664	Sequence 16664, A
C 32	151.5	4.0	663	4	US-09-252-991A-30843	Sequence 30843, A
C 33	151	4.0	226	4	US-09-252-991A-20432	Sequence 20432, A
C 34	150.5	4.0	297	4	US-09-252-991A-28842	Sequence 28842, A
C 35	150.5	4.1	529	4	US-09-252-991A-18245	Sequence 18245, A
C 36	150	4.0	256	4	US-09-252-991A-23974	Sequence 23974, A
C 37	150	4.1	267	4	US-09-252-991A-28620	Sequence 28620, A
C 38	150	4.1	591	4	US-09-252-991A-28760	Sequence 28760, A
C 39	149.5	4.0	681	4	US-09-252-991A-24567	Sequence 24567, A
C 40	149	4.0	319	4	US-09-252-991A-32635	Sequence 32635, A
C 41	149	4.0	442	4	US-09-252-991A-23285	Sequence 23285, A
C 42	148.5	4.0	432	4	US-09-252-991A-23622	Sequence 23622, A
C 43	147.5	3.9	375	4	US-09-252-991A-24278	Sequence 24278, A
C 44	147.5	4.0	433	4	US-09-252-991A-28695	Sequence 28695, A
C 45	147.5	4.0	491	4	US-09-252-991A-31758	Sequence 31758, A

ALIGNMENTS

RESULT 1
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6285192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Alignment Scores:
Pred. No.: 3,15e-191 Length: 386
Score: 2038.00 Matches: 386
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 54.46% Indels: 0
DB: 3 Gaps: 0

US-09-645-078-1 (1-2043) x US-09-045-284A-2 (1-386)

QY	128	ATGCTACTGCTAAATAATGAAGTCCCTGCTTCCAGATGCCATCTTG	187
Db	1	MeLeuLeuProLysLysMetLysLeuLeuLeuLeuLeuLeu 20	
QY	188	GCCTCTATTCTCCACATGTACGCCACACATCAGCTCCCTCTCTATGAGCAGGCC	247

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2038	54.5	386	3 US-09-045-284A-2
2	2038	54.5	386	4 US-09-190-911-1
3	1984	53.0	386	4 US-09-786-240-11
4	598.5	16.0	483	3 US-09-263-023-2
5	598.5	16.0	483	4 US-09-471-867-2
6	587.5	15.7	484	3 US-09-263-023-4
7	587.5	15.7	484	4 US-09-471-867-4
8	549	14.7	411	4 US-09-015-188-2
9	500.5	13.4	458	2 US-08-655-878-2
10	482	12.9	479	2 US-08-899-514-2
C 11	184	5.0	472	4 US-09-252-991A-31978
C 12	170	4.6	618	4 US-09-252-991A-23373

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 6, 2004, 10:57:23 ; Search time 24 Seconds
(without alignments)
8789.322 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 3742

Sequence: 1 gaattccattgtgttgga.....tgggattctataaaaaaaa 2043

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 389414 seqs, 51625971 residues.

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODSL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09645078/runat_06052004_104600_21782/app_query.fasta_1.2183
-DB=Issued Patents AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFT=PCT -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09645078.ecgn_1_1_27@runat_06052004_104600_21782 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PTCUT.COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*


```
QY 728 GTGACCTGTCCGGGACCCCGGCGCTGTTCCGTTCCGAGACGACAAAGGAGAT 787
Db 201 ValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAsp 220
QY 788 CTATGATTGACAGTCGATGCGGCGAGCATGAGCAAAACTCAAGAAAGGAGGAC 847
Db 221 LeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysGluAsp 240
QY 848 CAACCTACTATGTAGTCAGGTCATCTGCCAAAGCCAGCTGAGATCTCAAGACATC 907
Db 241 GlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
QY 908 CAGTCCTTGCCCAAGGCGCTGCAGAAACGCTACCTGCTTGTGGCTATGAGGACCTGGCT 967
Db 261 GlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla 280
QY 968 CGAGCCCTGTGGCCAGACTTCCCGAATGTATGAATTCGTGGATTTGGAATCTTGCCC 1027
Db 281 ArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPro 300
QY 1028 CATCTTCAGACCTGGGTGCATACATCACCGAGGCAAGGCGCATGGGTGACCAAGCTTTC 1087
Db 301 HisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
QY 1088 CACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGGTCTTTGCCCTAT 1147
Db 321 HisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTyr 340
QY 1148 GAAAGGTTTCTGACATTCAGAAAGCTGTGGCGATGCCATGAATTTGCTGGGCTACCGC 1207
Db 341 GluLysValSerArgLeuGlnLysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrArg 360
QY 1208 CACGTCAGATCTGAACAGAACAGAGAAACCTGTGTGTCGATCTTCTGTACCTGGAAT 1267
Db 361 HisValArgSerGluGlnGluGlnArgAsnLeuLeuLeuAspLeuLeuSerThrTrpThr 380
QY 1268 GTCCCTGAGCAATCCAC 1285
Db 381 ValProGluGlnIleHis 386

RESULT 3
; US-09-786-240-11
; Sequence 11, Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11

Alignment Scores:
Pred. No.: 6,29e-186 Length: 386
Score: 1984.00 Matches: 383
Percent Similarity: 99.48% Conservative: 2
Best Local Similarity: 98.97% Mismatches: 1
Query Match: 53.02% Indels: 2
DB: 4 Gaps: 0
US-09-645-078-1 (1-2043) x US-09-786-240-11 (1-386)
QY 128 ATGCTACTGCTTAAAAAATGAAGCTCCTGCTGTTCTGGTTTCCAGATGCCATCTTG 187
Db 1 MetLeuLeuProLysLysMetLysLeuLeuLeuPheLeuValSerGlnMetAlaIleLeu 20
QY 188 GCTCTATTCTTCCACATGTACAGCCACAAATCAGCTCCCTGTCTATGAAGGCCACGCC 247
Db 21 AlaLeuPhePheHisMetTyrSerHisAsnIleSerSerLeuSerMetLysAlaGlnPro 40
QY 248 GAGCCATGCGAGCTGCTGTTCTGCTTCTTCTGGGCGCTCTGGCTCTTCTTTTGTGGGCG 307
Db 41 GluArgMetHisValLeuValLeuSerSerTrpArgSerGlySerSerPheValGlyGln 60
QY 308 CTTTTTGGGCGAGCACCCAGATGTTTCTACCTGATGGAGCCGCTGCGACGTGTGATG 367
Db 61 LeuPheGlyGlnHisProAspValPheTyrLeuMetGluProAlaTrpHisValTrpMet 80
QY 368 ACCTTCAAGCAGAGCACCCGCTGCTGTCATGCTGTCATGCTGCGGATCTCATAGGCC 427
Db 81 ThrPheLysGlnSerThrAlaTrpMetLeuHisMetAlaValArgAspLeuIleArgAla 100
QY 428 GTCTTCTTGTGGACATCAGAGCGTCTTTCATGCTACATGGAACTGGTCCCGGAGACAG 487
Db 101 ValPheLeuCysAspMetSerValPheAspAlaTyrMetGluProGlyProArgArgGln 120
QY 489 TCCAGCCTCTTTCAGTGGGAGAACCGCGGCGCTGTTCTGTCACCTGCTGTGACATC 547
Db 121 SerSerLeuPheGlnTrpGluAsnSerArgAlaLeuCysSerAlaProAlaCysAspIle 140
QY 548 ATCCACAGATGAATCATCTCCCGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 607
Db 141 IleProGlnAspGlu-SerSerProGlyLeuThrAlaGlySerCysAlaValAsnSerPr 160
QY 608 TTTGAGGTG-GTGAGAAAGGCTGCGCTCTCTACAGCCACGCTGCTGCTCAAGAGGTGCG 666
Db 160 OLeuLysLeuLeuGluLysAlaCysArgSerTyrSerHisValValLeuLysGluValArg 180
QY 667 CTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726
Db 180 gPhePheAsnLeuGlnSerLeuTyrProLeuLysAspProSerLeuAsnLeuHisIle 200
QY 727 CGTGCACTGTGTCGGGACCCCGGCGCTGTTCCGTTCCGAGAACGACGACAAAGGAGGAGA 786
Db 200 eValHisLeuValArgAspProArgAlaValPheArgSerArgGluArgThrLysGlyAs 220
QY 787 TCTCATGATTGACATGCGATTTGATGGGCGAGCATGAGCAAAACTCAAGAGGAGGA 846
Db 220 pLeuMetIleAspSerArgIleValMetGlyGlnHisGluGlnLysLeuLysLysGluAs 240
QY 847 CCAACCTACTATGTAGTCAGGTCATCTGCCAAAGCCAGCTGAGATCTCAAGACCAT 906
Db 240 pGlnProTyrTyrValMetGlnValIleCysGlnSerGlnLeuGluIleTyrLysThrIle 260
QY 907 CCAGTCTTGTGCCAAGGCGCTGCGAGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 966
Db 260 eGlnSerLeuProLysAlaLeuGlnGluArgTyrLeuLeuValArgTyrGluAspLeuAla 280
QY 967 TCGAGCCCTGTCGGCCAGACTTCCCGAATGTATGAATTTGCTGGATTTGGAATCTTTCGCC 1026
Db 280 aArgAlaProValAlaGlnThrSerArgMetTyrGluPheValGlyLeuGluPheLeuPr 300
QY 1027 CCATCTTCAGACCTGGGTGCATACATCACCGAGGCAAGGCGCATGGGTGACCAAGCTTT 1086
Db 300 oHisLeuGlnThrTrpValHisAsnIleThrArgGlyLysGlyMetGlyAspHisAlaPhe 320
```

```

Qy 1087 CCACACAAATGCCAGGGATCCCTTAATGCTCCAGGCTTGCGGCTGTGCTTGGCCCTA 1146
Db 320 eHisThrAsnAlaArgAspAlaLeuAsnValSerGlnAlaTrpArgTrpSerLeuProTy 340
Qy 1147 TGAAGAAGTTTCTCGACTTCAGAAAGCTCTGGCGATGCCATGAATTTGCTGGGCTACCG 1206
Db 340 rGluIysValSerArgLeuGlnIysAlaCysGlyAspAlaMetAsnLeuLeuGlyTyrAr 360
Qy 1207 CCACGTCAGATCTGAACAAGACAGAGAAACCTGTGTGCTGGATCTTCTGTACCTGGAC 1266
Db 360 gHisValArgSerGluGlnGluArgAsnLeuLeuAspLeuLeuSerThrTrpTh 380
Qy 1267 TGTCCCTGAGCAAAATCCAC 1285
Db 380 rValProGluGlnIleHis 386

RESULT 4
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLMUCOSAMINE-6-O-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/09/263.023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

Alignment Scores:
Pred. No.: 6,81e-50 Length: 483
Score: 598.50 Matches: 134
Percent Similarity: 54.40% Conservative: 64
Best Local Similarity: 36.81% Mismatches: 137
Query Match: 15.99% Indels: 29
DB: Gaps: 6

US-09-645-078-1 (1-2043) x US-09-263-023-2 (1-483)
Qy 248 GAGCGCATGCAGCGCTGGTCTCTCTTCGGGCTCTGGCTCTTCTTTTGGGGCAG 307
Db 116 LysArgGlnLeuValTyrValPheThrThrTrpArgSerGlySerSerPheGlyGlu 135
Qy 308 CTTTGTGGGACACCCAGATGTTTCTACCTGATGGAGCCGCGCTGGCAGCTGTGGATG 367
Db 136 LeuPheAsnGlnAsnProGluValPhePheLeuTyrGluProValTrpHisValTrpGln 155
Qy 368 ACCTTCAAGCAGACGCGCTGTGATGCTGCACATGCTGCGGGATCTGATACGGGCC 427
Db 156 LysLeuTyrProGlyAspAlaValSerLeuGlnGlyAlaAlaArgAspMetLeuSerAla 175
Qy 428 GTCCTTCTGTGCGACATGAGCGTCTTTTGATCGCTACATGGAACCTGGTCCCGGAGCAG 487
Db 176 LeuTyrArgCysAspLeuSerValPheGlnLeuTyrSerProAlaGlySerGlyGlyArg 195
Qy 488 TCC-----AGCCTCTTTCAGTGGGAGAACCGCGGGCCCTGTGTCTGCACCT 535
Db 196 AsnLeuThrThrLeuGlyIlePheGlyAlaAlaThrAsnLysValValCysSerPro 215

```

```
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

Alignment Scores:
Pred. No.:      6,818-50      Length:      483
Score:          598.50       Matches:     134
Percent Similarity: 54.40%   Conservative: 64
Best Local Similarity: 36.81% Mismatches:    137
Query Match:      15.93%     Indels:      29
DB:              4         Gaps:        6

US-09-645-078-1 (1-2043) x US-09-471-867-2 (1-483)

QY 248 GAGCGATGACGTGGTCTTGTCCTCCCTGGCGCTCTGCCTCTTTTGTGGGGCAG 307
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 116 LysArgGlnLeuValTyVaIPheThrTrpArgSerGlySerPhePheGlyGlu 135
QY 308 CTTTTTGGCACACGACGATGTTTCTACCTGATGAGCGCGCTGGACGTCGTGATG 367
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 136 LeuPheAsnGlnAsnProGluValPhePheLeuTyrgluProValTrpHisValTrpGln 155
QY 368 ACCTTCAAGCAGACGCCCTGGATGCTGCACATGGCTGTGGGATCTGATACGGGCG 427
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 156 LysLeuTyrrProGlyAspAlaValSerLeuGlnGlyAlaalaArgaspMetLeuSerAla 175
QY 428 GTCTTCTTGGCACATGAGCGTCTTTGATGCCTTACATGGAACCTGTCTCCCGAGACAG 487
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 176 LeuTyrrArgCysAspLeuSerValPheGlnLeuTyrrSerProAlaGlySerGlyArg 195
QY 488 TCC-----AGCCTCTTTTCAGTGGGAGAACAGCGGGCGCTGCTTCTGCACCT 535
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 196 AsnLeuThrrLeuGlyIlePheGlyAlaalaThrAsnLysValValCysSerSerPro 215
QY 536 GCCTGTGACATCATCCACAAGATAATCATCCCCGGGCTCACTCAGGCTCTGTG--- 592
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 216 LeuCysProAlaTyrr--ArglysGluValValGlyLeuValAspAspArgValCysLys 234
QY 593 ---TGCAGTCAACAGCCCTTTGAGTGTGTGGAGAAGGCTCCGCTCCTACAGCCACGTG 649
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 235 LysCysProGlnArgLeuAlaArgPheGluGluGluLucysArgLysTyrrAargThrVal 254
QY 650 GTCTCAAGGAGTGGCGCTTCTTAACCTGCAGTCCCTCTACCGCGCTGCTGAAGACCCC 709
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 255 ValIleLysGlyValArgValPheAspValAlaValLeuAlaProLeuLeuLysAspPro 274
QY 710 TCCTCAACTGATATCTGTGACCTGTGTCGGGACCCCGCGGCGGTCCGCTTCCTCCGA 769
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 275 AlaLeuAspLysValIleHisLeuValArgaspProArgAlaValAlaSerSerArg 294
QY 770 GAACGCACAAAGGAGATCTCATGATTACATGTCGATTGTGATG-----814
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 295 IleArgSerArgHisGlyLeuIleArgGluSerLeuGlnValValArgSerArgaspPro 314
QY 815 -----GGGCAGCATGAGCAAAAACCTCAACAAGGAG 844
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 315 ArgAlaHisArgMetProPheLeuGluAlaAlaGlyHisLysLeuGlyAlaLysLysGlu 334
QY 845 GACCAA-----CCCTACTAT-----GTGATGCAGGTCATCTGCCAAGC 893
    ::::| | | | | ::::| | | | | ::::| | | | | ::::| | | | |
Db 335 GlyMetGlyGlyProAlaAspTyrrHisAlaLeuGlyAlaMetGluValIleCysAsnSer 354
QY 884 CAGCTGGAGATCTACAAAGACCATCCAGTCTTTGCCCAAGGCCCTTCGAGGAACGCTACTG 943
```

Db	355	MetAlaLysThrLeuGlnThrAlaLeuGlnProPAspTTrpLeuGlnGlyHisTyrLeu	374
Qy	944	CTTGTGGGCTATGAGGACCTGGCTCGAGCCCTGTGGCCGAGACTTCCCGAATGATGAA	1003
Db	375	ValValAlaGlyTyrGluAspLeuValGlyAspProValLysThrLeuArgValTyrAsp	394
Qy	1004	TTCTGGGATTGGAAATCTTCCCCCATCTTCAGACCTGGGTGCATAACATCACCCGAGGC	1063
Db	395	PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly	414
Qy	1064	AAGGCGATGGTGACCGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCTCCGAG	1123
Db	415	SerGlySerSerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn	434
Qy	1124	GCTTGGCCCTGGTCTTTCCTCCCTATGAAAGGTTTCTCGACTTCAGAAAGCCCTGTGCGGAT	1183
Db	435	AlaTrpAlaGlyThrAlaLeuThrPheGlnGlnLeuLysGlnValGluGluPheCysTyrGln	454
Qy	1184	GCCATGAATTTGTGGGCTACGGCACCTCGATCTGAAACAAGACACAGAAACCTGTTG	1243
Db	455	ProMetAlaValLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer	474
Qy	1244	CTGGATCTTCTG	1255
Db	475	LysThrLeuLeu	478

RESULT 6

US-09-263-023-4

; Sequence 4, Application US/09263023

; Patent No. 6037159

; GENERAL INFORMATION:

; APPLICANT: Uchimura, Kenji

; APPLICANT: Muramatsu, Hideki

; APPLICANT: Kadomatsu, Kenji

; APPLICANT: Kannagi, Reiji

; APPLICANT: Habuchi, Osami

; APPLICANT: Muramatsu, Takashi

; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

; TITLE OF INVENTION: DNA ENCODING THE SAME

; FILE REFERENCE: TOYAMA1, 001AUS

; CURRENT APPLICATION NUMBER: US/09/263,023

; CURRENT FILING DATE: 1999-03-05

; EARLIER APPLICATION NUMBER: JP 10-54007

; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: JP 10-177844

; EARLIER FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 484

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-263-023-4

Alignment Scores:	8,19e-49	Length:	484
Pred. No.:	587.50	Matches:	131
Score:	54.4%	Conservative:	67
Percent Similarity:	35.9%	Mismatches:	137
Best Local Similarity:	15.70%	Indels:	29
Query Match:		Gaps:	6
DR:			3

US-09-645-078-1 (1-2043) x US-09-263-023-4 (1-484)

Qy	248	GAGCGCATGACAGTGCCTGGTTCTCTCTCTCTGTCTCTGGCGCTCTGCTCTCTCTTTTGTTGGCGCAG	307
Db	117	LysArgHisTrpMetTyValPheThrTrpArgSerGlySerPhePheGlyGlu	136
Qy	308	CTTTTTGGCGACACCACCCAGATGTTTCTACTGATGAGCCCCGCCCTGGCAGTCGTGTGGATG	367
Db	137	LeuPheAsnGlnAsnPrGluValPhePheLeuTyrgLubroValTrpHisValTrpGln	156

296	IleArgSerArgHisGlyLeuIleArgGluSerLeuGlnValValArgSerArgAspPro	315
815	-----GGCGACGATGAGCAAAACATCAAGAAGGAG	844
316	ArgAlaHisArgMetProPheLeuGluAlaAlaGlyHisIleuGlyAlaLysLysGlu	335
845	-----GACCAACCCCTACTAT-----GTGATGCGAGGTCACTGCCAAAGC	883
336	GlyValGlyGlyProAlaAspTyrHisAlaLeuGlyAlaMetGluValIleCysAsnSer	355
884	CAGCTGGAGATCTACAGACCATCCAGTCTTCCCAAGGCCCTGCAGGAAACCTACCTG	943
356	MetAlaLysThrLeuGlnThrAlaLeuGlnProProAspThrPheGlnGlyHisTyrLeu	375
944	CTTGTCGGCTATCAGGACCTGGCTCGAGGCCCTGTGGCCCGACATCTCCCGAAATGTATGAA	1003
376	ValValArgTyrGluAspLeuValGlyAspProValLysThrLeuArgArgValTyrAsp	395
1004	TTGCTGGGATTGGAAATCTTGCCCATCTCTCAGACCTGGGTGCATACATCACCCGAGGC	1063
396	PheValGlyLeuLeuValSerProGluMetGluGlnPheAlaLeuAsnMetThrSerGly	415
1064	AAAGGATGGTGACACCGCTTCCACAAATGCCAGGATGCCCTTAATGTCTCCAG	1123
416	SerGlySerSerSerLysProPheValValSerAlaArgAsnAlaThrGlnAlaAlaAsn	435
1124	GCTTGGCGTGTGCTTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAGCCTGTGGCGAT	1183
436	AlaTrpArgThrAlaLeuThrPheGlnGlnIleLysGlnValGluGluPheCysTyrGln	455
1184	GCCATGAATTGTGTGGCTACCGCCACGTCAGATCTCAAACAAGACAGAGAAACCTCTTG	1243
456	ProMetAlaValLeuGlyTyrGluArgValAsnSerProGluGluValLysAspLeuSer	475
1244	CTGGATCTTCTG	1255
476	LysThrLeuLeu	479

RESULT 8

US-09-015-188-2
; Sequence 2, Application US/09015188C

; Patent No. 6399358

GENERAL INFORMATION:

: APPLICANT: Williams, Kevin J

APPLICANT: Tabas, Ira

; TITLE OF INVENTION: A Human-Computer Interface for the Control of a Robot Arm

INVENTION: 6-Sulfotri-

FILE REFERENCE: JEFF-0231

: CURRENT APPLICATION NUMBER: 1

CURRENT FILING DATE: 1998-01-

CURRENT FILING DATE: 12-30-01
: NUMBER OF SEO ID NOS: 17

NUMBER OF SEQ ID NOS: 1;
: SOFTWARE: PatentIn Ver. 2.0

: SEC ID NO 2

: SEQ ID NO 2
: LENGTH: 411

LENGTH: 411
TYPE: PRT

LIFE: FNT
ORGANISM: Homo sapiens

US-09-015-188-2

2-POST-CO-20

Alignment Scores:

Alignment Scores:

FILED: NO.:	SCORE:
4:32E-17	549.00

SCOUTS: 50
percent similarity.

PERCENT SIMILARITY:	50.12%
Best Local Similarity:	34.32%

BEST LOCAL SIMILARITY: 34.32%

Query Match: 14.6/8

DB: 4

115-00-245-050-1 (1-3003) ~ 0511-0511-0

US-09-645-078-1 (1-2043) X US-0

E
C
E
C
E
C
C
K
K
U
E
K
K
K
K

C
C

QY 143 AAAATGAAGCTCCTGCT

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

bb 6 LysAlaValLeuLeuLeu

(C)
K
U
A
E
U
E
A
C
K
U
U
E
U
E
E

A
C
T

:
(C)

194 TTCTCCACATGTACAG
QY

Db 378 GlnValLeuAlaGlnLeuGlyTyrLysIleAlaAlaSerGluGluGluLeuLysAsnPro 397

Qy 1241 TTGCTGATCTCTG 1255

Db 398 SerValSerLeuVal 402

RESULT 9

US-08-655-878-2

; Sequence 2, Application US/08655878

; Patent No. 5927713

; GENERAL INFORMATION:

; APPLICANT: FUKUTA, MASAKAZU

; APPLICANT: HABUCHI, OSAMI

; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE:

; STREET:

; CITY:

; STATE:

; COUNTRY:

; ZIP:

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/655,878

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME:

; REGISTRATION NUMBER:

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE:

; TELEFAX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 458

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-655-878-2

Alignment Scores:

Pred. No.: 2,758-40 Length: 458

Score: 500.50 Matches: 120

Percent Similarity: 54.02% Conservative: 75

Best Local Similarity: 33.24% Mismatches: 133

Query Match: 13.38% Indels: 33

DB: 2 Gaps: 13

US-09-645-078-1 (1-2043) x US-08-655-878-2 (1-458)

Qy 227 CTGCTATGAGGACACCCGAGCGGATGACGTGCTGTTCTGCTTCTGCGGCTCT 286

Db 104 LeuGlyIleAlaAlaProGluProArgHisValLeuLeuMetAlaThrThrArgThr 123

Qy 287 GCCTCTCTTTGTCGGGACGCTTTTGGGAGCAGCCAGATGTTTCTACCTCATGGAG 346

Db 124 GlySerSerPheValGlyGluPheAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 143

Qy 347 CCGGCTGGCAGTG- - - - - TGGATGACCTTCAAGCAGACACCGGCTGGATGCTGAC 400

Db 144 ProLeuTrpHisIleGluArgThrValThrPheGluProGlyGlyAlaAsnAlaValGly 163

Qy 401 ATGGCTGTG- - - - - CCGGATCTGATACGGCCCTCTTCTGTGCGATGAGGCTTT 454

Db 154 SerAlaLeuValTyrArgAspValLeuGlnGlnLeuLeuLeuCysAspLeuTyrIleLeu 183

Qy 455 CATGCTACATGGAACCTGTCCTCCGAGACAG- - - - - TCCAGCCTCTTCTAGTGGAG 508

Db 184 GluSerPheIleSerProAlaProGluHisLeuThrAlaAlaLeuPheArgGly 203

Qy 509 AACAGCGGCGCTGCTGTCGACCTGTCGACATCATCCACAA- - - - - GATGAATC 565

Db 204 SerSerHisSerLeuCysGluGluProValCys- - - - - ThrProSerLeuLysVal 221

Qy 566 ATCCCGGCTCATCTGAGG- - - - - CTCTGTGAGTCAACAGCCCTTTGAGGTGGAG 622

Db 222 PheGluLysTyrHisCysLysAsnArgCysGlyProLeuAsnIleThrLeuAlaAla 241

Qy 623 AAGGCTGCTGCTCTCAGCAGCGTGTCTCAAGGAGGTGCGCTTCTTCAACCTGAG 682

Db 242 GluAlaCysArgArgLysGlnHisMetAlaLeuLysThrValArgIleArgGlnLeuGlu 261

Qy 683 TCCTCTACCCGCTGCTGAAGACCCCTCCCTCAACCTGCATATGTCGACCTGTCGG 742

Db 262 PheLeuGlnProLeuAlaGluAspProArgLeuAspLeuArgIleIleGlnLeuValArg 281

Qy 743 GACCCCGGCGGTGTCGCTCCGAGAGCGCAAGGAGATCTCATGATTGACAGT 802

Db 282 AspProArgAlaValLeuValSerArg- - - - - MetVal- - - - - 292

Qy 803 CGCATTGTGATGGGCGCAGCATGACGAAAAA- - - - - TCAAG- - - - - 838

Db 293 - - - - - AlaPheSerGlyLysTyrGluSerTrpLysTyrPheAlaGluGlyGluAlaPro 311

Qy 839 - - - - - AAGGAGGACCAACCTACTATGTGTGATGAGTCACTGCCAAGC- - - - - CAGCTGGAG 892

Db 312 LeuGlnGluAspGlu- - - - - ValGlnArgLeuArgGlyAsnGlySerIleArgLeuSer 330

Qy 893 ATCTACAAGACCATCCAGTCTTCCCAAGGCGCTGCGAGAACGCTACCTCTTGTGCGC 952

Db 331 AlaGluLeuGlyLeuArgGln- - - - - ProArgTrpLeuArgGlyArgTyrMetLeuValArg 349

Qy 953 TATGAGGACCTGCTGAGCCCTGTCGCCAGACTTCCCGAATGTATGATTCGTGGGA 1012

Db 350 TyrGluAspValAlaArgAlaProLeuArgLysAlaLeuGluMetTyrArgPheAlaGly 369

Qy 1013 TTGGAATCTTCCCTCATCTTCCAGACCTGGTGTGATACATCATCCCGAGGCAAGGCGCATG 1072

Db 370 IleHisProThrProGlnValGluGluTrpIleArgAlaAsnThrGlnAlaPro- - - - - Gln 388

Qy 1073 GGTGACCAAGCTTCCACAAATCCAGGAGTCCCTTAATGTCTCCAGCGCTGGCGC 1132

Db 389 AspSerAsnGlyIleTyrSerThrGlnLysAsnSerSerGluGlnPheGluLysTrpArg 408

Qy 1133 TGTCTTTGCTTATGAAAGGTTTCTGACTTCAGAAAGCTGTGGCGCATGCAATGAAT 1192

Db 409 PheSerIleProPheLysLeuAlaGlnValGlnAspAlaCysGluProAlaMetArg 428

Qy 1193 TTGCTGGGTACCGCCAGCTGATCTGAACAGAACAGAGAAACCTCTTGTGATCTT 1252

Db 429 LeuPheGlyTyrLysLeuAlaSerSerAlaGlnGluLeuThrAsnArgSerLeuSerLeu 448

Qy 1253 CTG 1255

Db 449 Leu 449

RESULT 10

US-08-655-878-2

; Sequence 2, Application US/08899514

; Patent No. 5910581

; GENERAL INFORMATION:

; APPLICANT: HABUCHI, OSAMI

; APPLICANT: FUKUTA, MASAKAZU

; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN

; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING

; TITLE OF INVENTION: FOR THE POLYPEPTIDE

; NUMBER OF SEQUENCES: 9

Thu May 6 11:30:32 2004

CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
CITY: NEWPORT BEACH
STATE: CALIFORNIA
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

NAME: DANIEL E ALTMAN
REGISTRATION NUMBER: 34

REGISTRATION NUMBER: 34, 113
REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
TELECOMMUNICATION INFORMATION:

TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502

FORMATION FOR SEQ ID NO: 2
SEQUENCE CHARACTERISTICS:
LENGTH: 479

LENGTH: 475
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
-899-514-2

ment Scores: 1 040-30

```
NO.: 1.84E-38
: 482.00
nt Similarity: 52.72%
```

Local Similarity: 32.38%
Match: 12.88%

DB:	2	Gaps:
US-09-645-078-1	(1-2043)	x US-08-899-514-2 (1-479)

QY 251 CGCATGCACGTCGTGGTTCGTCTTCCTGGCGCTCTGGCTCTTCTTTTGTGGGACGCT 310

Db
132 ArgArgHisValLeuLeuMetAlaThrThrArgThrGlySerSerPheValGlyGluPhe 151

[illegible]

DO	152	FREASNGINGINGLYASHALLEPNEIYLEUFRNEGRIUOLEUFRPHLSIREGUARHNI I/
QV	371	-----TTCAAGCAGAGACCCGCTGGATGCTGCACATGGCTGTG-----CGGGATCTG 418

Db 172 ValSerPheGluProGlyClyAlaAsnAlaAlaClySerAlaLeuValTyrArgAspVal 191

QY 419 A T A C G G C G C T C T T C T T G T G C G A C A T G A C G G T C T T T G A T G C C T A C A T G G A A C C T G G T C C C 478

Db 192 LeuLysGlnLeuPheLeuCysAspLeuTyrValLeuGluHisPheIleThrProLeuPro 211

QY 479 CGGAGACAGTCCAGCCTCTTTTCACTGGAG-----AACAGCCGGGCCCTGTGTCTGCA 532

	GluAspHisLeuthrGlnPheMet-PheArgArgGlySerSerArgSerLeuCysGlu	Asp 231
D _b	212 GluAspHisLeuthrGlnPheMet-PheArgArgGlySerSerArgSerLeuCysGlu	
C _a	532 GAGGGCGTCCTGTACATCATTCCCAAGAATCATTTTGCCTGCTTGCTGCC	589

QY CGCGCCGAGACATCATCCCCAAGAGAGAAATCACTCCCCGGGCATCATTGAGG--CTC 389
||| ||| : : : : : : : : : : : :
Db CCGCCGAGACATCATCCCCAAGAGAGAAATCACTCCCCGGGCATCATTGAGG--CTC 390

232 ProValCysThrPropheValLys--LysValPhegluIvsTvrHisCysLysAsnArg 250
||| ||| : : : : : : : : : : : :

QY 590 CTGTGAGTCAACAGCCCTTTGAGGTGGTGGAGAGGCCCTGCCGCTCCTACAGCCACGTG 649

Db 251 ArgCysGlyProLeuAnValThrLeuAlaAlaGluAlaCysArgArgLysGluHisMet 270

QY 650 GTGCTCAAGGAGGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGCTGAAGACCCC 709

271	AlaLeuLysAlaValArgIleArgGlnLeuGluPheLeuGlnProLeuAlaGluAspPro	290
710	TCCCTCAACTGCATATACGTGCACCTGGTCCGGGACCCCGGGCGGTGTTCGGTTC	769
291	ArgLeuAspLeuArgValIleGlnLeuValArgAspProArgAlaValLeuAlaSerArg	310
770	-----GAAAGCGCAAAAGGAGAGATCTCATGATTGACAGTCGCATTGTG	811
311	MetValAlaPheAlaGlyLysTyrLysThrTTPLeuAspAspGluGlyGln	330
812	ATGGGGCAGCATGAGCAAAACTCAAGAAGGAGGACCAACCTACTATGTGATCAGGTC	871
331	AspGlyLeuArgGluGluGluValGlnArg-----LeuArgGly	343
872	ATCTGCCAAAGC--CAGCTGGAGATCTACAAGACCATCCAGTCTCTGCCCAAGCCCTG	928
344	AsnCysGluSerIleArgLeuSerAlaGluLeuGlyLeuArgGln---ProAlaTrpLeu	362
929	CAGGAAGCGTCACTGCTTGTGGCTATGAGGACCTGGTCCAGGCCCTGGTGGCCCGAC	989
363	ArgGlyArgTyrMetLeuValArgTyrGluAspValAlaArgGlyProLeuGlnLysAla	382
989	TCCCGAATGATGAATTCGTGGGATGGGAATTCCTGGCCCATCTTCAGACCTGGGTGCAT	1048
383	ArgGluMetTyrProPheAlaGlyIleProLeuThrProGlnValGluAspTrpIleGln	402
1049	AACATCACCCGAGCGAAGCGCATGGGTGACCAACGCTTCCACACAAATGCCAGGATGCC	1108
403	LysAsnThrGlnAlaAlaHisAspGlySerGlyIleTyrSerTrpGln---LysAsnSer	421
1109	CTTAATGTCCTCCAGGCTTGGCGCTGGTCTTTGCCCTATGAAGGTTTCTCGACTTCAG	1168
422	SerGluGlnPheGluLysTrpArgPheSerMetProPheLysLeuAlaGlnValGln	441
1169	AAAGCCTGGCCATGCCATGAATTTGCTGGGCTACCCGACGTCCAGATCTGTAACAAGAA	1228
442	AlaProCysGlyProAlaMetArgLeuPheGlyTyrLysLeuAlaArgAspAlaAla	461
1229	CAGAGAAACCTGTTGCTGATCTTCTG	1255
462	LeuThrAsnArgSerValSerLeuLeu	470

RESULT 11

```

US-09-252-991A-31978
; Sequence 31978, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31978
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31978

```

Alignment Scores:	3.24e-09	Length:	472
Pred. No.:	Score:	Matches:	115
	184.00	Conservative:	372
Percent Similarity:	31.7%	Mismatches:	172
Best Local Similarity:	24.84%	Indels:	144
Query Match:	4.97%	Gaps:	23
DB:	4		

DB: 4

US-09-645-078-1 (1-2043) x US-09-252-991A-31978 (1-472)

QY 1142 GCAAGACCAAGCGCCAGGCTGGGAGACATTAAAGGCATCCCTGGCATTTGTGTGGAAG 1083
 Db 55 AlaProArgGlyAlaGlyProGlyProGlyAlaArgPro----- 68
 QY 1082 CGTGGTCACCATCCCTTGCCTCGGTGATGTTATGACCCAGGCTGGAAGTGGGCA 1023
 Db 69 ArgGlyAspArgAlaAspArgAlaGlySerAlaThrSerProArgAlaPro----- 86
 QY 1022 AGAATTCATCCACGAATTCATACATTCGGGAAGTCTGGG----- 981
 Db 87 -----ProGlyProSerProHisHisThrProGlyValProGlyAlaAspAlaThr 104
 QY 980 -----CCACAGGGCTCGACCGAGTCTCTCATAGCCACAGCAGTACGGTT 933
 Db 105 ProGlyAlaAlaProGlySerValThrValArgAlaHisArgGlySerAlaAspAlaGly 124
 QY 932 CCTCAGGCGCT-----TGGGCAAGGACTGGA-----TGGTCT 900
 Db 125 ThrAlaArgProArgArgHisThrAlaAlaThrAlaValArgSerAlaArgHisAla 144
 QY 999 TGATATCTCCAGTGGCTTTGGCAGATGACCTGCATCATAGTAGGTTGGTCTCTCT 840
 Db 145 AlaAlaArgProAlaAlaAlaLeuArgGlnProAlaProHisSerGlyGlySerProPro 164
 QY 839 TCTTGAGTTTGTCTGCTGCTGCCCATCATCAATGCGACTGTCAATCATGAGATCTCCCT 780
 Db 165 ThrSerAlaLeuArgHis-----ArgArgValArgCys-----Pro 176
 QY 779 TTGTGCGTTCGGAACGGA-----ACAGGCGCGGGGTCCCGGA----- 738
 Db 177 GlyAlaAlaAlaGlyAsnGlyArgArgAlaAlaArgProGlyArgProGlyGlySerPro 196
 QY 737 ---CCAGGTGCACATATGAGGT---TGAGGGAGGGTCTTTCAGCAGCGGTAGAGGG 684
 Db 197 LeuAlaGlyGlnGluArgAlaGlyValValAlaProAlaValGlnThrAlaProProAla 216
 QY 683 ACTGAGGTTGAAGAAGCGCACCTCTTGTAGCACCAGCTGGCTGTAGAGCGGCGAGG--- 627
 Db 217 ThrAlaGly-----AlaProAlaGlyTrpArgHisArgArgAlaGlyArgPro 233
 QY 626 -----CCTTCTCCACCTCAAGGGGCTGTGACTGCACAGGA 588
 Db 234 AlaSerArgLysArgProProAlaProProProArg----- 246
 QY 587 GCCTGACGTAGCGCGGGGATGATTCATCTTGGGATGATGTACAGCAGGTGCAG 528
 Db 247 -----CysArgArgThrValArg 252
 QY 527 AACACAGGCGCGGCTGTCTCCACAGAGAGGCTGAGTGTCTCCGGGACGAGTT 468
 Db 253 ArgSerThrProGlyAlaProAlaAlaHisProGlyGluProProAlaAlaProProIle 272
 QY 467 -----CCATGTAGGCAT-----CAAGACGCTCA 444
 Db 273 ArgProGlyProValArgArgValProProAlaThrArgProAlaSerAspArgArgAla 292
 QY 443 TGTGCG-----ACAAGAAGCGCGGCTATCAGATCCCGCACAG--- 405
 Db 293 IleArgArgThrProAlaGlySerValArgArgProProAlaLysArgGlyGlnSer 312
 QY 404 -----CCATGTGACGCA-----TCCAGGCGG 384
 Db 313 ProAlaProAlaAlaValProAlaGlyAlaAlaProAlaLeuArgArgArgAla 332
 QY 383 TGCTCTGTGAAGTCTACACAGTCCAGGCGGCTCCATCAGT---AGAAACAT 327
 Db 333 AlaThrAlaAlaAlaAspSerArgLeuAlaArgAlaProAlaGlyArgArgArgHis 352
 QY 326 CTGGGTGTCGCCAAAGAGCTGCCCAAAAGAGACCGACGAGCGCC----- 279
 Db 353 AlaAlaProAlaSerProAlaArgProValProAlaAlaLeuAlaAlaGlyAsn 372

QY 278 AGAAGACAGAACACAGCAGCTGCATCGCTGGGCTGTGCTTCATAGACA----- 228
 Db 373 CysArgArgAspSerGlyArgHisGlyIleArgAlaGlyProAlaGlyThrHisArgSer 392
 QY 227 ---GGGAGCTATGTTGGCTGTACATGTGAAGAATAGAGCAAGATGGCATCTGGG 171
 Db 393 ProGlyProGlyCysArgGlyAlaArgGlnArgAsnAlaAlaGlyGlyCysProAlaGly 412
 QY 170 AAACCAAGAACACAGCAGGAGCTTCATTTTTTAGGCAGTAGCATTTGCTGAAGTGAAGA 111
 Db 413 LysProArgAlaSerAlaAlaSerIle-----Arg 422
 QY 110 CTTTGTGCTGGGCTTCCCTGCTGCTTTTCAGAGGCTCACTGTAGCGGCGCATGGC 51
 Db 423 ValCysSerSerAlaGluArgAlaArgTyzSerAlaSerSerThrSerArgSerArg 442
 QY 50 TTTTCT-----CCTCTA 39
 Db 443 CysAlaSerMetSerSerLysGlyArgAlaSerMetSerLysSerArgAlaArgProAla 462
 QY 38 CCCTTCTCT 30
 Db 463 ProMetCys 465

RESULT 12

US-09-252-991A-23373
 ; Sequence 23373, Application US/09252991A
 ; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 23373
 LENGTH: 618
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23373

Alignment Scores:
 Pred. No.: 8,82e-08 Length: 618
 Score: 170.00 Matches: 133
 Percent Similarity: 32.88% Conservative: 35
 Best Local Similarity: 26.03% Mismatches: 191
 Query Match: 4.59% Indels: 152
 DB: 4 Gaps: 29

US-09-645-078-1 (1-2043) x US-09-252-991A-23373 (1-618)

QY 1585 CTTCCAGGCGCCAAAGTCTCTGAAGTCTGCCACAGAGGATCAAGAAAGAGAGA 1526
 Db 158 ProProThrHis-----ArgArgArg 165
 QY 1525 GCAATACCTGTTTCACTGCTGGG----- 1502
 Db 166 AspLeuProAlaAlaThrGlyArgHisTyProGlyArgLeuGlnLeuGlyProGlyGly 185
 QY 1501 -----ATGTGCTGCATCAAGTTCCTCCAGTCTGTTTCTAGACAAGATGGACA 1448
 Db 186 GlnPheArg-ProProGlnLeuArgLeuProLeuLeuSerArg-----GlyHis 202
 QY 1447 CAAAAGTCTTCTGCTTGAGCAGCTGTGCACACAACTCACACTCAGTGTGATCCACAG 1388
 Db 202 salaglyProProGly---AlaLeuLeuArgProGlyHis---HisLeuAspGlnAlaAr 220

Db 82 -----AlaAlaGlyProIleProArgAlaArg----- 90
Qy 701 TCACACGGGTAGAGGAGTCAAGTGTGAAGAGCGCACCTCTTTCAGCACACGTCGGC 642
Db 91 GlyProAlaProIleGlyArgGlyAlaProSerAlaProPro---AlaProGlyTrp 109
Qy 641 TGTAGGAGCGGCGCTCTTCCACCACTCAAGGCTGTGACTGC-----ACA 591
Db 110 SerArgAlaArgCysProAlaProProGlyCysAlaThrSerAlaArgSerArgArg 129
Qy 590 GGACCTGCTAGTACCGGGGGATGATTTTCATCTTGTGGATGATGTCACAGCAGGTG 531
Db 130 GlyAlaAlaAlaAenProArg-----CysGlyArgGlyThr 141
Qy 530 -----CAGAACAACAGG-----CCGCGCTCTTCTCCCACT 501
Db 142 ProProArgSerArgGlyThrValArgArgValArgArgCysProGlyCysGlyAlaAla 161
Qy 500 GAA-----AGAGGCTGACTGTCTCCGGGACAGGTTCCATGATGATGATCAAGA 450
Db 162 GluSerArgProAlaArgTrpGlyArgSerAspSerMetGlyArgArgSerGlySer 181
Qy 449 CGCTCATGTCGACAGA-----AGACGGCCCGTATCAGATCCCGACAG-----CCA 402
Db 182 ArgArgSerArgIleProGlyCysArgSerValAlaAlaAlaProAlaGluSerArgPro 201
Qy 401 TGTGCA-----GCATCCAGCGGTCTCTCTGAAGTTCATCCACAGTCCAGCGG 348
Db 202 CysThrArgTrpSerGlyArgCysSerValAlaArgProAlaGlyArgGlyArg 221
Qy 347 GCTCCA---TCAGGTAGAAACATCTGGTCTGCCAAAAGTGTCCCAACAAAG 291
Db 222 ArgProArgSerGlyArg-----CysProGlyArgPro 232
Qy 290 AGC-----CAGAGCGCCAGAGACAGAACACAGACGTGTGATCGTGGGTGTGCCT 237
Db 233 SerSerValArgAsnAlaArgCysProSerProGlyArgArgValAla----- 248
Qy 236 TCATAGACAGGAGCTGATGTGTGCTGATCTGGAAGATAGAGCCAGAGTGCCA 177
Db 249 -----AlaCysArgArgThrAlaArgArgGlyArg 258
Qy 176 TCTGGGAACCAAGAACACAGAGCTTCTATTTTATTTAGGAGTATGTCGAAGT 117
Db 259 Ser-----Ser 260
Qy 116 GGAAGACCTTGTGGC 102
Db 261 CysArgProCysGly 265

RESULT 14
US-09-252-991A-23948
; Sequence 23948, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23948
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23948
Alignment Scores:

Pred. No.: 3,49e-07 Length: 335
Score: 162.50 Matches: 92
Percent Similarity: 32.39% Conservativity: 34
Best Local Similarity: 23.65% Mismatches: 124
Query Match: 4.34% Indels: 139
DB: 4 Gaps: 25
US-09-645-078-1 (1-2043) x US-09-252-991A-23948 (1-335)
Qy 12 TGTGGGTACGAAGACACACAGAGGTAGAGGAGAAAGCGCATGCGCGGTAGCAGT 71
Db 43 CysTrpArgGluSerThr-----SerThrGln 52
Qy 72 GAGCCTCTCAAGACAGACAGGAGGAGCCACAGCCACAGGTTCTCCACTTCAGCAATGC 131
Db 53 ArgSerThrSerSerThrTrpArgAlaSerThrArgSerAlaThrSerProAlaCys 72
Qy 132 TACTGCTTAAAAA---TGAAGCTCTGCTGTCTTCTGGTTCCAGATGCCCATCTGG 188
Db 73 ArgSerAlaArgArgCysArgThrGlyCysCys-----ArgArgProThrArg 88
Qy 189 CTCTATTCTTCACATGT-----ACAGCCACACATCAGCT 224
Db 89 SerSerSerSerThrCysArgAlaSerCysTrpSerAlaCysAlaArgAlaArgSer 108
Qy 225 CCTGTCTATGAAGCAC-----AGCCCG 248
Db 109 ThrCysArgAsnArgArgGlyArgSerMetArgSerSerArgArgProThrSerPro 128
Qy 249 AGCGATGACAGTGTCTGCTGCTCTCTGCGCGCTGCGCTCTCTCTTTTGGGGCAGC 308
Db 129 ---ArgCysValAsnTrpArgCysArgProProPro----- 139
Qy 309 TTTTGGGCGACACAGATGTTTCTACCTGATGGAGCGCGCTGGACGCTGTGGATGA 368
Db 140 ---ArgGlySerThr-----ProThrSerThrAlaThr 150
Qy 369 CTTTCAAGC-----AGACACCGCTGATGCTGCACATGCTGCGGGATC 416
Db 151 AlaSerAlaAlaArgGlyArgArgCysAlaGlyCysTrpTrpAlaSerThrVal 170
Qy 417 TGATACGGCGCTCTCTCT---TGTCGACATGAGCGTCTTTGATCGCTACATGGAACCTG 473
Db 171 ThrAsnArgProAsnAlaTrpCysAlaThrProAlaGly---SerProAsnValGlyThr 189
Qy 474 GTCCCGGGA-----GACAGTCAGCTCTTTTCAGTGGGAGAACACCGCGGCC 521
Db 190 CysProGlyAlaTrpTrpThrSerIleProAlaAlaCysAlaAlaSerThrArgGly--- 208
Qy 522 TGTGTTCTGCACCTGCTGTGACATCATCCACAGATGAATCATCC---CCCGGGCTC 578
Db 209 CysAlaCys-----ArgValArgSerAsnTrpProSerGly 220
Qy 579 ACTGCAGGCTCTGTGTCAGTCAACAGCCCTTTGAGGTGTGGAGAGGCGCT----- 629
Db 221 SerAlaAlaArgSerSerCysCysAlaProAlaArg---TrpArgGlyProSerCysAsn 239
Qy 630 -----CCCGCTCTCAGCCAGCGTGTCTCAAGGAGTCCCTCTTCTCAACC 677
Db 240 MetProAlaSerAlaAlaProAlaTrpSerTrpLeuAlaSerArgGlyAlaAlaGlyThr 259
Qy 678 TGCAGTCCCTCTTACCCGCTGCTGAAAGACCCCTCCCTCAACCTGCATATCGTGACCTGG 737
Db 260 AlaGlyProSerAlaValVal----- 266
Qy 738 TCCGGGACCCCGGGCGCTGTTCCGTTCCGAGAAAGCAAGAGGAGATCTCATGATTG 797
Db 267 -----SerProGlyValCysCysAlaThrAlaArgAlaTrpArgSerValPro----- 282
Qy 798 ACAGTGCATTGTGATGGGCGCATGAGCAAAAAGTCAAGAGAGGAGGACCAACCCCTACT 857
Db 283 -----TrpThrAlaArgProThrCysAlaArgArgAsnArgGlyHisAla 297

Thu May 6 11:30:32 2004

Db 418 laGlyArgArgArgGluAspArgArgSerThrArgProSerAlaAlaAlaGln 437

Search completed: May 6, 2004, 11:08:13
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:47:57 ; Search time 20 Seconds
(without alignments)
1856.496 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKXKMLFLVQMAIL.....EQRNLLDLLSTWTVPQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623	30.6	484	2 JC7350	N-acetylglucosamin
2	606.5	29.8	486	2 JC7351	N-acetylglucosamin
3	588.5	28.9	484	2 JE0261	N-acetylglucosamin
4	500.5	24.6	458	2 A57397	chondroitin 6-sulf
5	168	8.2	307	2 E95934	probable enzyme, C
6	116	5.7	388	2 G70729	hypothetical prote
7	95	4.7	358	2 I40485	surfactin synthet
8	92.5	4.5	2354	2 T13288	mei-41 protein - f
9	92	4.5	359	2 T16350	hypothetical prote
10	91.5	4.5	338	2 JE0196	hydroxysteroid sul
11	91	4.5	659	2 S11737	resistance protein
12	91	4.5	1847	2 E64477	replication factor
13	89.5	4.4	346	2 E86319	probable flavonol
14	89.5	4.4	504	2 T07120	probable cytochrom
15	89.5	4.4	4273	2 C69579	polyketide synth
16	89	4.4	524	1 A36205	mitochondrial proc
17	89	4.4	1140	2 G71335	probable transcript
18	89	4.4	1968	1 PN0093	genome polyprotein
19	88	4.3	359	2 T47187	hypothetical prote
20	88	4.3	466	2 E84132	aminopeptidase B3
21	87.5	4.3	380	2 T42755	tyrosylprotein sul
22	87.5	4.3	4063	2 T42993	probable spectrin
23	87.5	4.3	4101	2 T23630	hypothetical prote
24	87	4.3	2500	2 G88493	protein F57B9.2 [i
25	86.5	4.2	748	1 S08680	methyilmalonyl-CoA
26	86.5	4.2	831	2 S26675	DNA-directed DNA p
27	86.5	4.2	2352	2 T06077	splicing factor PR
28	86	4.2	320	2 A40216	flavonol 4'-sulfo
29	86	4.2	823	2 S48986	probable protein k

30 86 4.2 869 2 A96558 probable protein k
31 85.5 4.2 344 2 H83928 hypothetical prote
32 85.5 4.2 764 1 S14113 1-phosphatidylinos
33 85 4.2 632 2 T18692 hypothetical prote
34 85 4.2 672 2 T34899 probable DNA recom
35 85 4.2 1237 2 T37529 hypothetical prote
36 85 4.2 2688 2 I49477 alpha-A-crystallin
37 84.5 4.1 217 2 JQ1358 Ci protein - Misc
38 84.5 4.1 646 2 G86863 threonine-trNA lig
39 84.5 4.1 1031 2 T43458 hypothetical prote
40 84.5 4.1 2493 2 S45734 probable membrane
41 84 4.1 390 2 AB2337 group 2 sigma 70-t
42 84 4.1 1189 2 T42726 guanine nucleotide
43 84 4.1 1217 2 T00270 hypothetical prote
44 84 4.1 1583 2 T14176 probable phosphati
45 83.5 4.1 518 2 E84536 hypothetical prote

ALIGNMENTS

RESULT 1

JC7350

N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C:Species: Mus musculus (house mouse)

C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000

C:Accession: JC7350

R:Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.

Biochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a

A:Reference number: JC7350; PMID:20374462; PMID:10913333

A:Accession: JC7350

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB040710

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylat

sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand 1

C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 30.6%; Score 623; DB 2; Length 484;

Best Local Similarity 34.2%; Pred. No. 1.5e+45; Indels 46; Gaps 7;

Matches 136; Conservative 65; Mismatches 151;

QY	30	NISSLSKKAQPERMHVLVLSWSRSGSFVGLQFQHPDVFYLMPEPAHVMWTFKQSTAWM	89
DB	87	NLSAVGEAVTQEQHIVVHATWRTGSSFLGELFQHPDVFYLYEPMMHLQALYPGDAES	146
QY	90	LHMAVRDLIRAVFLCDMSVFDAYMEPGPRQ-----SSLFOWENSRALCSAPACD	139
DB	147	LQALRDMRLSLPRCDFSLRLYAQPGDGERAPDSANLTTAMLFRTWTKVICSPLCP	206
QY	140	IIPQDE----IIPRAHCLRLCSQOPFEVVEKACRSYSHVVLKVEYFFNLOSFLPKDPS	195
DB	207	AAPRARADVGLVEDKACESTCPPVSLRALAECHKYPVWIKDVLRLDLGLVPLLRDPG	266
QY	196	LNLIHIVHLVRPRAVFRSRETKGDLMDS-----RIVM-----CQHE	233
DB	267	LNKVVQLFRDPRAVHNSLRKQGLRESIQVLRTRQGRGHFRHVLVLAHGVDPGQA	326
QY	234	OKLKKEQDPY---VMQVICQSQLEIKYKITSQSLPKALQERYILVRYEDLARAFVAQTSR	289
DB	337	RALSAFADFLTSALEVICEAWLRDLITFRGAPAWLRRYLRLRYEDLVWQQAQLRR	386
QY	290	MYEFVGLFEFLPHLOTWVHNITRGKMG-DIAFHTNARDALNVSOAWTWSLPEYKVSLOK	348
DB	367	LIRFESGLTALAALDAFAFNMTGRSAYGADRPFLSARDAREAVHVRERLSQEQVQVET	446
QY	349	ACGDAMNLLGYRHSVEQEQENLL-----LDLLSTWTW	381
DB	447	ACAPAMRLIATPRSGDENDRXTVREGTPTETKANAV	484

RESULT 2

JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; transmembrane protein

Query Match 29.8%; Score 606.5; DB 2; Length 486;
Best Local Similarity 37.0%; Pred. No. 3.9e-44;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERMHVLSSWRSGSFVGGQFPGHPDVFYLMPEAWHVMWTFKQSTAWMLHMAVRDLIRA 100
DB 100 EKQHIYVHATWRTGSGFLGELFNQNPVEFFLYEPVHWQKLYPGDAVSLQGAARDMLSA 176
QY 101 VFLCMSVFDAYMEPG-PRRQ-----SSLFQWNSRALCSAPACDIIPDEIIIPRAHCRLL- 155
DB 177 LYRCDLSVFLQYSPAGSGRNLTLTGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDDRCK 235
QY 156 -CSQOPFEVVEKACRSYSHVVLKEVRFENLQSLVPLKDPNSLNHLHIVHLVRDPRVAFRSR 214
DB 236 KCPORLARFECECKRYTLNKGVRVDFVAVLAPLLRDPALDLKVIHLVRDPRVAFRSR 295
QY 215 ERTKGLMIDSRIVM-----GQHEOKLKE-----DQPY-----VMQVICS 252
DB 296 IRSRGLRESIQVRSRDPRAHRMFFLEAAHGKLGAKKEGVGGPADYHALGAMEVICNS 355
QY 253 QLEIKYTIOSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELPHLOTWVHNIRG 312
DB 356 MAKTUQTALQPPDWQGHVLYVRYEDLVGBPVKTRRVIDFVGLLVSEMEQFALNMTSG 415
QY 313 KGMGDHAFHTNARDALNVSQAWNSLPYKVSRLQKACGDAMNLLGYRHVRSEQQRNLL 372
DB 416 SGSSSKFFVVSARNATQAANAWRTALTQQIKQVEFCYQPMVILGYRVRVNSPEEVKOLS 475
QY 373 LDLL 376
DB 476 KTL 479

RESULT 4
A57397
Chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C:Species: Gallus gallus (Chicken)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
C:Accession: A57397
R;Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi,
J. Biol. Chem. 270, 18575-18580, 1995
A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransf
A:Reference number: A57397; MUID:95355490; PMID:7629189
A:Accession: A57397
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-458 <FUK>
A:Cross-references: GB:D49915; NID:G971262; PIDN:BAA08655.1; PID:G971263
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match 24.6%; Score 500.5; DB 2; Length 458;
Best Local Similarity 33.2%; Pred. No. 4.6e-35;
Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

QY 34 LSMKAQPERMHWLVLSWSRSGSFVGGQFPGHPDVFYLMPEAWHV--WMTFKQSTAWMLH 91
DB 104 LGTAAPRPRHULLMATTGSGFVGEFFNQGNIFLYFEPLWHIERVTTFEPGGANAVG 163
QY 92 MAV--RDLIRAVFLCMSVFDAYMEPGPRRQ--SSLFQWNSRALCSAPACDIIPQ-DEI 146
DB 164 SALVYRDVLQQLLDCDLYLESFISPAEHLTAALFRGSGSHSLCEEVPC--TPSLKKV 221
QY 147 IPRAHCR-LLCSQOPFEVVEKACRSYSHVVLKEVRFENLQSLVPLLKDPNSLNHLHIVHLVR 205
DB 222 FEKHCKNRRCGPLNITLAAECRRKQHWALKTVIRIQLEFLOPLAEDPLDLRIIQLVR 281
QY 206 DPRVAFRSRERTKGLMIDSRIVMVGQHEOKL-----KEDQPYVMQVICS-QUE 255
DB 282 DPRVAVLSR-----MV---APSGYESKWKAAEAGEAPLOEDE-VORLGNCSIRLS 330
QY 256 IYKTIOSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELPHLOTWVHNIRGKM 315
DB 331 AEGLGKQ-PWLRGRVWLVRYEDVAPRKPALKALEMTFRFAGIHTPTQVEWIRANTQAP-Q 388
QY 316 GDHAFHTNARDALNVSQAWNSLPYKVSRLQKACGDAMNLLGYRHVRSEQQRNLLDL 375

JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; transmembrane protein

Query Match 29.8%; Score 606.5; DB 2; Length 486;
Best Local Similarity 37.0%; Pred. No. 3.9e-44;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERMHVLSSWRSGSFVGGQFPGHPDVFYLMPEAWHVMWTFKQSTAWMLHMAVRDLIRA 100
DB 100 EKQHIYVHATWRTGSGFLGELFNQNPVEFFLYEPVHWQKLYPGDAVSLQGAARDMLRS 159
QY 101 VFLCMSVFDAYMEPG-PRRQ-----SSLFQWNSRALCSAPACDIIPDEIIIPRAHCRLL- 146
DB 160 LFRCDFSVRLYAPGDPAAAPDPTANLTALFRWTRTKVICSPPPLCPGAPARAEEVGL 219
QY 147 IPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVRFENLQSLVPLLKDPNSLNHLHIVHLVRD 206
DB 220 VEDTACERSCPPVAITALEAECKYPVVVKDRLDLGLVPLLRDPGLNKLKVQLFRD 279
QY 207 PRAVFRSERTKGLMIDSRIVM-----GQHEOKLKEEDQPY 244
DB 280 PRAVHNSRLKSRQGLRESIQVLRTRQGRDFRHFVLLAHGVGARPGGSRALPAAPRADF 339
QY 245 ----VMQVICSQSLBIYKTIOSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELP 300
DB 340 FLTGALVCEAWLRLDFARCAPAWLRVRLRYEDLVQPRALRLRLRFSGLRALA 399
QY 301 HLQTWVHNITRGKMGV-DHAFHTNARDALNVSQAWNSLPYKVSRLQKACGDAMNLLGY 359
DB 400 ALDAPALANTRGAAYGADRPFLSARDAREAVHAWRRLSREQVRQVEACAPAMRLIAY 459
QY 360 RHRVSEQE 367
DB 460 ---PRSGEE 465

RESULT 3
JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C:Accession: JE0261
R;Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Pan, Q.W.; Mitsuoka,
T.
J. Biochem. 124, 670-678, 1998
A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
A:Reference number: JE0261; MUID:98391845; PMID:9722682
A:Accession: JE0261
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DDBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfate from 3'-phosphoadenos
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match 28.9%; Score 588.5; DB 2; Length 484;
Best Local Similarity 36.0%; Pred. No. 1.4e-42;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

Db 389 DNGIYSTOKNSSEFEKWRFSIPFKLAQVDDQADPAPNRLFGYKLAASAGELNRSLSL 448
QY 376 L 376
Db 449 L 449

RESULT 5
E95934
Probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sinc
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95934
R:Finan, T.M.; Weidner, S.; Wong, K.; Rühmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95934
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <KUR>
A:Cross-references: GB:AL591385; PIDN:CAC49141.1; PID:gl5140626; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb21237
A:Genome: plasmid

Query Match 8.2%; Score 168; DB 2; Length 307;
Best Local Similarity 22.7%; Pred. No. 9.2e-07;
Matches 83; Conservative 50; Mismatches 131; Indels 102; Gaps 16;

QY 36 MKAQPERMHVLSSWRSGSFVQGLFGQHPDVF---YLMPEAPHWMTFKQSTAWMLHM 92
Db 1 MFSQVRI-AYTAGYGRSGSTILDALGQHAAVGVANGITSLTRVWRHNEYCA----- 53
QY 93 AVRDILRAVFLCDMSVFDAYMEPPGRQSSLFQWNSRALCSAPACDIPODEIIPRAHC 152
Db 54 -----CGNAIRDCSFWSVRR-----EWSDG-----QDPGLMBEYC 84
QY 153 RLICSOQPPEVVEKACR-----SYSHVVLKEVRFPN-LQS----- 186
Db 85 AL---QKKEGLSMVTRLLSGMLGKQFSLYLHTKRLFSAMQSCSGRQIVDSKILPR 141
QY 187 LYPLLKPSLNLIHVLVDPDPAVFRS-----RERTKGLMIDSRIVMGQHEQKKEKD 240
Db 142 ANAVAQIPGIDMRVTHLVDRGRGVAWSLLKGVYERDAKSG-----LQKEI 185
QY 241 QPYVMQVTCQSQLEIYKTIQSLPAL-QERYLLVRYEDLAPAPVAQTSRMVYFVGLFL 299
Db 186 KPKSVFRTALRWSM-VNLAVEYLSRKLGSKEKVMRYVEDFASDPVAVMQQIGTFLFD-L 243
QY 300 PHLQTVWHNITRGKMG-----DHAFHTNARDALNVSQAWRWSLPYKVSRLQKACGDA 353
Db 244 SQVGTSLN-GEAMGPGHGVAGNRLRMVASIALNKDFTWTRMPARQVQSFQLGGWM 300
QY 354 MNLGY 359
Db 301 LRRYG 306

RESULT 6
G70729
hypothetical protein RV2267c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70729
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70729
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-388 <COL>
A:Cross-references: GB:Z77163; GB:AL123456; NID:G3261610; PIDN:CAB00968.1; PID:e255072;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2267c

Query Match 5.7%; Score 116; DB 2; Length 389;
Best Local Similarity 19.3%; Pred. No. 0.037;
Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

QY 45 VLVLSWRSGSFVQGLF---GQH--PDVFLMEPAPHWMTFKQSTAWMLHVAVDLIR 99
Db 83 IFIVGHWRGTGTTLLHLLVVDNRHTGTGYECLAP--HHPLL-----TEW----- 125
QY 100 AVFLCDMSVFDAYMEPPGRQSSLFQWNSRALCSAPACDIIPQ-DEII-----PRAHC 152
Db 126 -----FAPVVE-----FLVSKHRANDMDLSLHHPQEDFVCMQGLPSPYL 167
QY 153 RLICSOQP-----FEVFKAKSYSHVVLKEVRFPNLSLY-----PLLKDPSLNLHI 200
Db 168 TIAPFNRPPQYEEYLDLEQVAPRELE--IWKTLFRFVQCVYFRRKTVILKNPHTSFRI 225
QY 201 -----VHLVRDPRAVFRSERTKGLMIDSRIVMGQHEQKKEQDPYY--VM 246
Db 226 KVLEVPQAKFHIVRDPVYV-----PSTHLKALYRIHGLQOPTDGLD 273
QY 247 QVTCQSQLEIYKTIQSLPALQ-ERYLLVRYEDLAPAPVAQTSRMVYFVGLF-----FLPH 301
Db 274 DKVSTVVDILYRLKDEGRELVDPTRFVELYEDLIGDPGQLRLYQHLGLGDFECLPLR 333
QY 302 LQTVWHNITRGKMGDHA-PHTNARDALNVSQAWRWSLPYKVSRLQKACGDMNLGY- 359
Db 334 LRQI-----LADHADYKINS-----YQLTVQRAIVDEHGEIIDRYGVD 373
QY 360 RH 361
Db 374 RH 375

RESULT 7
I40485
Surfactin synthetase component I - Bacillus subtilis
N:Alternate names: competence protein srfAA; surfactin production protein srfAA; surfaci
C:Species: Bacillus subtilis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000
C:Accession: I40485; B63718; S35517; A37323; S46967; A43705; S34985
R:Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van Sin
Mol. Microbiol. 8, 821-831, 1993
A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthesis
A:Reference number: I40485; MUID:93360813; PMID:8355609
A:Accession: I40485
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3588 <RES>
A:Cross-references: EMBL:X70356; NID:G396480; PIDN:CAA9816.1; PID:G396481
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

QY 340 YEKVSRLOKACGDAMNLLGYRHVRSQ-----EQRNLLDLLSTWTVPEQI 385
 DB 1344 KLLVSRASVNCGEVARALSYLESLEGEDSKORLLSQFTFLVVEVYGLRDPDSV 1398
 RESULT 9
 T16350
 hypothetical protein F42G9.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
 C:Accession: T16350
 R:Taich, A.
 submitted to the EMBL Data Library, March 1996
 A:Description: The sequence of C. elegans cosmid F42G9.
 A:Reference number: Z18498
 A:Accession: T16350
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-359 <TAI>
 A:Cross-references: EMBL:U00051; NID:G1216305; PID:G1216308; PIDN:AAA91354.1; CESP:F42G9
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F42G9.8
 A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
 C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8
 Query Match 4.5%; Score 92; DB 2; Length 359;
 Best Local Similarity 19.5%; Pred. No. 3.8; Mismatches 114; Indels 98; Gaps 15;
 Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;
 QY 52 RGSSEVQLFCQHPDV----FYLMEPAWHVMTFKQSTAMNLMHMAVRDLIRAVFLCDMS 107
 DB 95 RSGTLMRAILDAPDVRCGGTMLPSFLTQAGWRND-WYNSGI-----TQE 143
 QY 108 VPDAYMEPPRQSSFLQWNSRALCSAPACDIIPQ-DEIIPRAHCRLLCSQOPFEVVEK 166
 DB 144 VDDAV-----SAFTEIVAKHSELAPR-----LCKNDP----- 172
 QY 167 ACRSYSHVVLKEVRFNFLQSLPLKDPNLHIVHLVRDPRAVRSRRTK----GDLN 222
 DB 173 ----YTALWLPTR-----RLYP-----NAKFLIMRIDARAVVHSMIEKVPVAGYNT 216
 QY 223 IDSRVIMGQHEOKLKEQDPYVVMQVICOQ-----LEIYKTIQSLPKALQERYLLVRYED 278
 DB 217 SDIEISFVQWNGELRK-----MTFCNNAPGQCIVY-----YER 251
 QY 279 LARAPVAQTSRMVYFVGLFELPHLQTVVHNITRGKMGDHAFH-TNARDALNVQAMRWS 337
 DB 252 LIQKPAEETLRITNFDLPFSQMLRHQDLIGDEVLDNQPSASQVKNSTKALTTSWF 311
 QY 338 LPY-EKVSLQACGDAMNLLGY 359
 DB 312 DCFSEETLRKLDVDVAPFLGILGY 334
 RESULT 10
 JE0196
 hydroxysteroid sulfotransferase (EC 2.8.-.-) - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 18-Jun-1999
 C:Accession: JE0196
 R:Sakakibara, Y.; Yanagisawa, K.; Takami, Y.; Nakayama, T.; Suiko, M.; Liu, M.C.
 Biochem. Biophys. Res. Commun. 247, 681-686, 1998
 A:Title: Molecular cloning, expression, and functional characterization of novel mouse
 A:Reference number: JE0196; MUID:98321187; PMID:9647753
 A:Accession: JE0196
 A:Molecule type: mRNA
 A:Residues: 1-338 <SAK>
 A:Cross-references: GB:AF026072; NID:G2570897; PIDN:AAAC69918.1; PID:G2570898
 C:Comment: This enzyme catalyzes the transfer of a sulfonate group from the active sulfa
 p.
 C:Superfamily: alcohol sulfotransferase
 C:Keywords: transferase

Query Match 4.5%; Score 91.5; DB 2; Length 338;
 Best Local Similarity 22.2%; Pred. No. 3.8; Mismatches 32; Indels 71; Gaps 11;
 Matches 52; Conservative 32; Mismatches 32; Indels 71; Gaps 11;
 QY 67 DVFYLMPEPAWHVMTFKQSTAMNLMHMAVRDLIRAVFLCDMSYFDAYMBPGPRRQSSLFQW 126
 DB 59 DIFIVTYP-----KSGTNWMB-----IVCLILKD-----GDPWSIRSPWQ- 96
 QY 127 ENSRALCSAPACD-IIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVRFENLQ 185
 DB 97 -----RAPWCETIISAFNVLDPRSPRIMSHLPIELFTKA-----FF--- 133
 QY 186 SLVPLKDPNLHIVHLVRDPRAVRSRRTKGDLMIDSRVIMGQHEOKLKEQDPYV 245
 DB 134 -----SSKAKVIYVGNPRDVVS-----LYYSKIA-GQ-----LKDPGTPDQF 172
 QY 246 MQVICQSQLEIYKTIQSLPKAL-----QERYLLVRVEDLARAPVAQTSRMVYFVG 295
 DB 173 LQNFLKGEVFGSFWFDHIKGMIRMQENFLFITVEELQQDLRGLSGVQRICEFLG 226
 RESULT 11
 S11737
 resistance protein Mx3, interferon-regulated - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Mar-2000
 C:Accession: S11737
 R:Meier, E.; Kunz, G.; Haller, O.; Arnheiter, H.
 submitted to the EMBL Data Library, April 1990
 A:Description: Unexpected activity of rat Mx proteins against a Rhabdovirus.
 A:Reference number: S11737
 A:Accession: S11737
 A:Molecule type: mRNA
 A:Residues: 1-659 <EMB>
 A:Cross-references: EMBL:X52713; NID:G56724; PIDN:CAA36937.1; PID:G56725
 C:Superfamily: dynamin-related protein VPS1

Query Match 4.5%; Score 91; DB 2; Length 659;
 Best Local Similarity 20.6%; Pred. No. 10; Mismatches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;
 Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;
 QY 62 FGQHPDVFYLMPEPAWHVMTFKQSTAMNLMHMAVRDLI-----RAVFLCDMSYFDAYME--- 114
 DB 301 FKHPQFRALLEDG-----KATVPCLAERLTMELISHICKSLPLENQIKESHQSTSE 353
 QY 115 -----PGPRRQSSFLQWNSRALCSAPACDIIPQDEIIPRAHCRLLCS-QQPF--- 161
 DB 354 ELQKYGADIPEDENEXTLELIEKINAF-NQDITAVEGEEIVREKECRLLTKLKEFFLW 412
 QY 162 -EVVEKACRSYSHVVLKEVRFENLQ-----SLYPLLKDPNLH 200
 DB 413 SEEIERNFQSGDALLYKEVYTFEMQYGRGLPGFVNYKTFENLIRQIKLSEPAEM--- 470
 QY 201 VHLVRD-PRAVPR-----SRRTKGDLMIDSRVIMGQHEOKLKEQDPYVVMQV 248
 DB 471 LHKVTEIVRAAFTVVSEKNFSEPFNLHRTTKSKLEDIRL---EQETEAKAIRLHFQMQ 527
 QY 249 ICOSQLEIYKTIQSLPKALQERYLLVRYEDL-----ARAPVAQTSRMVYFVG 296
 DB 528 IIVCQDQIVR-----KALQK-----VREBEEERKHKGRSAQSPNLTSSMD----- 572
 QY 297 EFLPHLQTV---VHN 308
 DB 573 EITQHLNAYRQEAHN 587
 RESULT 12
 B64477
 replication factor C homolog - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: B64477

QY	80 MTF-----KQTAWMLHMAVRDLIRAVFLCDMSVFDAYNEPQPRRQSLSFQW 126
----	--

Db 248 EDILRKH-----SEAEQEDLDVLL 279
EKTRVKEGNG-----

Search completed: May 6, 2004, 10:51:57
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:46:52 ; Search time 17 Seconds
(without alignments)
1182.299 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2039
Sequence: 1 MLPPKMKLLFLVSQMAIL.....EQRNLLDLLSTWTVPQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500.5	24.6	458	1 C6ST_CHICK	Q92179 gallus gall
2	116	5.7	388	1 YME7_MYCTU	Q50695 mycobacteri
3	104.5	5.1	419	1 NOE6_RHISN	P55472 rhizobium s
4	95	4.7	358	1 SRF1_BACSU	P27206 bacillus su
5	92	4.5	359	1 TP8B_CAEEL	Q20351 caenorhabdi
6	91.5	4.5	5430	1 MACF_HUMAN	Q9upn3 homo sapien
7	91.5	4.5	5938	1 MACF_HUMAN	Q96pk2 homo sapien
8	91	4.5	234	1 MX3_FAT	Q15777 homo sapien
9	91	4.5	639	1 MX3_FAT	P18590 rattus norv
10	89.5	4.4	504	1 C7D8_SOYBN	O81974 glycine max
11	89.5	4.4	4273	1 PKSM_BACSU	P40872 bacillus su
12	89	4.4	320	1 F4ST_FLACH	P52837 flaveria ch
13	89	4.4	524	1 MPPA_RAT	P20069 rattus norv
14	89	4.4	920	1 CAPP_STRPY	Q9a0u7 streptococ
15	89	4.4	1968	1 RRPO_PWR	P17965 porato viru
16	88	4.3	312	1 F3ST_FLABI	P52835 flaveria bi
17	88	4.3	2209	1 KNTC_HUMAN	P50748 homo sapien
18	87.5	4.3	380	1 TP8A_CAEEL	O77081 caenorhabdi
19	87	4.3	940	1 CHRD_BRARE	O57472 brachydanio
20	86.5	4.2	748	1 MUTA_MOUSE	P16332 mus musculu
21	86.5	4.2	831	1 DPOF_TRETH	P30113 thermus the
22	85	4.2	823	1 SCH9_YEAST	P11792 saccharomyc
23	85.5	4.2	559	1 BAL2_ARATH	Q9a8r4 arabidopsis
24	85	4.2	632	1 GLCE_CAEEL	P46555 caenorhabdi
25	85	4.2	932	1 CAPP_STRP3	Q8k873 streptococ
26	85	4.2	1237	1 YDY2_SCHPO	O13683 schizosacch
27	85	4.2	2688	1 ZEP1_MOUSE	Q30172 mus musculu
28	84.5	4.1	646	1 SVT_LACIA	Q9ced2 lactococcus
29	84.5	4.1	780	1 CUL5_MOUSE	Q9d5v5 mus musculu
30	84.5	4.1	2493	1 YEA4_YEAST	P35194 saccharomyc
31	83.5	4.1	803	1 MSHW_CANAL	Q94065 candida alb
32	83.5	4.1	864	1 GLND_PASMU	Q9cnd1 pasteurella
33	83	4.1	222	1 NOGG_XENLA	P49011 xenopus lae

ALIGNMENTS

RESULT 1

ID	C6ST_CHICK	STANDARD	PRT	458 AA
AC	Q92179;1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).			
OS	Gallus Gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
FN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.			
RC	STRAIN=White leghorn; TISSUE=Embryonic chondrocytes;			
RX	MEDLINE=95355490; PubMed=7629189;			
RA	Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,			
RA	Shinomura T., Habuchi O.;			
RT	"Molecular cloning and expression of chick chondrocyte chondroitin 6-			
RT	sulfotransferase.";			
RL	J. Biol. Chem. 270:18575-18580(1995).			
CC	-!- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN			
CC	SULFATE.			
CC	-!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin =			
CC	adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.			
CC	-!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By			
CC	similarity).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: D49915; BAA08655.1; ..			
DR	PIR: A57397; A57397.			
DR	InterPro: IPR000863; Sulfotransferase.			
DR	Pfam: PF00685; Sulfotransfer; 1.			
KW	Transferase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.			
FT	DOMAIN 1 23			
FT	TRANSMEM 24 37			
FT	DOMAIN 38 458			
FT	DOMAIN LUMENAL, CATALYTIC (POTENTIAL).			
FT	FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).			
SQ	SEQUENCE 458 AA; 52253 MW; C9A3B7D0A5086F0C CRC64;			
	Query Match 24.6%; Score 500.5; DB 1; Length 458;			
	Best Local Similarity 33.2%; Pred. No. 1.7e-34;			

Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;	
Qy	34 LSMKAQPERHVLVLSWSSGSPVGFQHPDPVFLMEPAHV--WMTFKQSTAWMLH 91
Db	104 LGIAAPRPRHLLMATRTGSGFVGFQGNIFLFEPLWHIERTVTFEGGANVG 163
Qy	92 MAV--RDLIRAVFLCDMSVFDAYMEPRRQ--SSLQWENSRLCSAPACDIPO-DEI 146
Db	164 SALVYRDVQLQLCDLVLLESFISPAPEEHLTAALFRGGSHSLCEEPVC--TPSLKKV 221
Qy	147 IPRAHOR-LLCSQOPFEWEKACSYSHVVLKEVRFENLQSLVPLKDPSLNLIHVLVR 205
Db	222 PEKTHCKNRGCPNITLAEACRKHMAKTVIRIQLEFLOPLADPLDLRIQLVR 281
Qy	206 DPRAVFRSRERTKGLMDSRVNMQHEQKILK-----KEDQPYVMQVTCOS-QLE 255
Db	282 DPRAVLVSR-----NV---APSGKYESMKMAEAGEAPLQDE-VQRLRGNCESIRLS 330
Qy	256 IYKTIQSLPKALQRYELLVRYEDLARAPVAQTSRMVEFVGLFELPHLQTVWHNITRGKM 315
Db	331 AELGLRQ-PRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHTPQVEEIRANTQAP-Q 388
Qy	316 GDHAFHTNARDALNVSOAWRSLPYEKVSRLOKACGDAMNLLGYRHRVRSQEORNLDDL 375
Db	389 DSNGIYSTQKNSSEQFEKWRISIFPKLAQVQDACEPAMRLFGYKLASSAQELTNRSL 448
Qy	376 L 376
Db	449 L 449
RESULT 2	
ID	YM67 MYCTU STANDARD; PRT; 388 AA.
AC	Q50695;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DE	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Hypothetical protein RV2267C/MT2329/Wb2290C.
GN	RV2267C OR M2329 OR MTCY339.43 OR MB2290C.
OS	Mycobacterium tuberculosis, and
OS	Mycobacterium bovis.
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1773, 1765;
RN	[1]_TaxID=1773, 1765;
SEQUENCE FROM N.A.	
RC	SPECIES=M.tuberculosis; STRAIN=H37RV;
RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares R.,
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RT	complete genome sequence."
RL	Nature 393:537-544(1998).
RN	[2]
SEQUENCE FROM N.A.	
RC	SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX	MEDLINE=22206494; PubMed=12218036;
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA	Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT	laboratory strains."
RL	J. Bacteriol. 184:5479-5490(2002).
RN	[3]

SEQUENCE FROM N.A.	
RC	SPECIES=M.bovis; STRAIN=AF2122/97;
RX	MEDLINE=22703107; PubMed=12788972;
RA	Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA	Prior M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA	Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA	Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT	"The complete genome sequence of Mycobacterium bovis."
RL	Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; 277163; CAB00968.1; ..
DR	EMBL; AE007076; AAK46611.1; ..
DR	EMBL; BX248342; CAD97151.1; ..
DR	FIR; G70729; G70729.
DR	TIGR; MT2329; ..
DR	Tuberculin; RV2267C; ..
DR	InterPro; IPR00863; Sulfotransferase.
DR	Pfam; PF00685; Sulfotransfer; 1.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 388 AA, 46062 MW, 5DED0263275A9B24 CRC64;
Query Match 5.7%; Score 116; DB 1; Length 388;	
Best Local Similarity 19.3%; Pred. No. 0.025;	
Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;	
Qy	45 VLVLSWSSGSPVGFQGLF---GQH--PDVFLMEPAHVWMTFKQSTAWMLHMAVRDLIR 99
Db	83 IFVGHWRITGLLHLLVDDRHGTGTCYECLAP--HFFLL---TEW----- 125
Qy	100 AVFLCDMSVFDAYMEPRRQSSLSFQWNSRLCSAPACDIPO-DEI-----PRAHC 152
Db	126 -----FAPYVE-----FLVSKHRAMNDLSLHHPQDEDFVWCQGLSPVYL 167
Qy	153 RLCSQOP-----FEVVEKACSYSHVVLKEVRFENLQSLY-----PLKDPSSLNLI 200
Db	168 TIAPFNRPPQBEYLDLEQVAPRELE--IWKTLPFRVQOVYFRRKVILKNPTHSFRI 225
Qy	201 -----VHLVRDPRAVFRSRERTKGLMDSRVNMQHEQKKEEDQPY--VM 246
Db	226 KVLEVPQAKPFIHVRDPVWV-----PSTHLKALYRIHGLQQTDFDGLD 273
Qy	247 QVTCQSLQSLYKTIQSLPKALQ-EXLLVRYEDLARAPVAQTSRMVEFVGLS-----FLPH 301
Db	274 DKVYSTVDLYRKLDEGRELVDPTRFYELKRYEDLIGDPGQLRLYQHLGLGDFCYLPR 333
Qy	302 LOTWVHNITRGKMGMDHA-FHTNARDALNVSOAWRSLPYEKVSRLOKACGDAMNLLGY- 359
Db	334 LRQY-----LADHADYKINS-----YQLTVEQRAIVDEHWHGEIIDRYGYD 373
Qy	360 RH 361
Db	374 RH 375
RESULT 3	
ID	NOEE RHISN STANDARD; PRT; 419 AA.
AC	P55472; P72326;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	Nodulation protein noe (EC 2.8.2.-).
GN	NOE OR Y4HS.
OS	Rhizobium sp. (strain NGR234).
OG	Plasmid sym pNGR234a.

RP SEQUENCE OF 1-460 FROM N.A.
 RX MEDLINE=91154134; PubMed=1847909;
 RA Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
 RA Zuber P.;
 RT "srfa is an operon required for surfactin production, competence
 development, and efficient sporulation in *Bacillus subtilis*.";
 RL J. Bacteriol. 173:1770-1778(1991).
 RN [6]
 RP SEQUENCE OF 1-64 FROM N.A.
 RC STRAIN=168;
 RA Fujishima Y., Yamane K.;
 RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
 of srfaA of the *Bacillus subtilis* chromosome.";
 RL Microbiology 141:277-279(1995).
 RN [7]
 RP SEQUENCE OF 1-38 FROM N.A.
 RX MEDLINE=91358326; PubMed=1715856;
 RA Nakano M.M., Xia L., Zuber P.;
 RT "Transcription initiation region of the srfa operon, which is
 controlled by the comp-comA signal transduction system in *Bacillus*
subtilis.";
 RL J. Bacteriol. 173:5487-5493(1991).
 CC -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -!- COFACTOR: Contains 3 covalently bound phosphopantetheines.
 CC -!- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
 CC -!- SIMILARITY: Belongs to the AMP-dependent AMP-binding enzyme
 CC family.
 CC -!- SIMILARITY: Contains 3 acyl carrier domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D13262; BAA02522.1; -;
 DR EMBL; X70356; CAA49816.1; -;
 DR EMBL; D50453; BAA08982.1; -;
 DR EMBL; Z99105; CAB12142.1; -;
 DR EMBL; M59939; AAA22815.1; -;
 DR EMBL; D30762; BAA21034.1; -;
 DR EMBL; M64702; AAA22816.1; -;
 DR PIR; I40485; I40485.
 DR HSP; P14687; 1AMU.
 DR Subtilast; BG10168; srfaA.
 DR InterPro; IPR00873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR006163; Pp bind.
 DR InterPro; IPR006162; Ppantne S.
 DR Pfam; PF00501; AMP-binding; 3.
 DR Pfam; PF00668; Condensation; 4.
 DR Pfam; PF00550; Pp-binding; 3.
 DR PRISM; PR00154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
 DR PROSITE; PS00455; AMP BINDING; 3.
 DR PROSITE; PS00075; ACP DOMAIN; 3.
 DR Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
 KW Multifunctional enzyme; Repeat; Complete proteome.
 FT REPEAT ? 1047 DOMAIN 1 (GLU-ACTIVATING).
 FT REPEAT ? 2084 DOMAIN 2 (LEU-ACTIVATING).
 FT REPEAT ? 3115 DOMAIN 3 (D-LEU-ACTIVATING).
 FT DOMAIN 976 1043 ACYL CARRIER (ACP) 1.
 FT DOMAIN 2013 2080 ACYL CARRIER (ACP) 2.
 FT DOMAIN 3044 3110 ACYL CARRIER (ACP) 3.
 FT BINDING 1006 1006 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT BINDING 2043 2043 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT BINDING 3074 3074 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT CONFLICT 146 146 M -> I (IN REF. 1 AND 5).

FT CONFLICT 151 151 V -> L (IN REF. 1 AND 5).
 FT CONFLICT 165 165 G -> A (IN REF. 1).
 FT CONFLICT 281 281 Q -> T (IN REF. 1 AND 5).
 FT CONFLICT 460 460 D -> T (IN REF. 5).
 FT CONFLICT 540 540 A -> P (IN REF. 1).
 FT CONFLICT 562 562 Y -> I (IN REF. 1).
 FT CONFLICT 639 639 Y -> GS (IN REF. 1).
 FT CONFLICT 644 644 P -> R (IN REF. 1).
 FT CONFLICT 647 649 FLP -> LLA (IN REF. 1).
 FT CONFLICT 1026 1026 Q -> L (IN REF. 1).
 FT CONFLICT 1065 1114 HWSHSEISSASSMRESAISASSNSSEGFYSKAGARIS
 INPTPRVIE -> SLAQRTYIVSQFEDAGVYNNPAAAIL
 EGPLDIOLKRAFOGLIRRHESLR (IN REF. 1).
 FT CONFLICT 1130 1131 VC -> DS (IN REF. 1).
 FT CONFLICT 1162 1162 A -> R (IN REF. 1).
 FT CONFLICT 1456 1456 L -> V (IN REF. 1).
 FT CONFLICT 1848 1848 L -> S (IN REF. 1).
 FT CONFLICT 1892 1892 T -> D (IN REF. 1).
 FT CONFLICT 1909 1909 A -> R (IN REF. 1).
 FT CONFLICT 1971 1978 PAVFIQMD -> LRCLSKWT (IN REF. 1).
 FT CONFLICT 2050 2050 P -> L (IN REF. 1).
 FT CONFLICT 2216 2216 Q -> H (IN REF. 1).
 FT CONFLICT 2263 2264 HR -> QQ (IN REF. 1).
 FT CONFLICT 2289 2289 L -> V (IN REF. 1).
 FT CONFLICT 2347 2347 C -> S (IN REF. 1).
 FT CONFLICT 2426 2426 S -> R (IN REF. 1).
 FT CONFLICT 2609 2610 AV -> RC (IN REF. 1).
 FT CONFLICT 2684 2686 MISSING (IN REF. 1).
 FT CONFLICT 2756 2756 S -> T (IN REF. 1).
 FT CONFLICT 2897 2898 TA -> SP (IN REF. 1).
 FT CONFLICT 3026 3026 N -> P (IN REF. 1).
 FT CONFLICT 3097 3097 N -> F (IN REF. 1).
 FT CONFLICT 3272 3272 S -> A (IN REF. 1).
 FT CONFLICT 3317 3317 S -> R (IN REF. 1).
 FT CONFLICT 3452 3452 S -> Y (IN REF. 1).
 FT CONFLICT 3484 3486 HQMSHPFFETISH -> DEMSDAGLFRSE (IN REF. 1).
 FT CONFLICT 3500 3501 PH -> GQ (IN REF. 1).
 FT CONFLICT 3508 3508 E -> R (IN REF. 1).
 SQ SEQUENCE 3588 AA; 402072 MW; BC02FBI57D7F1FDB CRC64;
 Query Match 4.7%; Score 95; DB 1; Length 3588;
 Best Local Similarity 19.7%; Pred. No. 24;
 Matches 63; Conservative 48; Mismatches 111; Indels 98; Gaps 13;
 QY 14 VSQMAILAFFMYSHINISLSMKAQPERMHVLSWSRSGSFVQQLFGQHPDVFYLM 73
 Db 990 VEKAGIFDNFFETGGHSLKXATLLTKHK-----ETGIEIPQQLFHEPTITALAE 1040
 QY 74 PAHWVMTFKQSTAWMLHMAVLDLIRAVFLCDMSVFDAYMEGPRROSSLFOWENSRALC 133
 Db 1041 EADH-----RESKAF-----AVIEPAKQEHYPLHWHSEHIS 1073
 QY 134 SAPA-----CDIIPQDEII-----PRAHCLLCS 157
 Db 1074 SASSRMRESAISASSNSSEGFYSKAGARISGINPTPRVIESFVLENSTPRQKIH-VCV 1132
 QY 158 QQPFVVEKACRSYSHVVLKEVRFNQLSLYLLKXDPISL-----NLHIV-----HLVRDR 208
 Db 1133 DFNIEIMERGRSDEAIMASFVRTDLAKA-PLFRIGLGLLENHMLLFDVHHLISDGV 1191
 QY 209 AVPRSEERTKGDLMID--SRIVMGQHEQKLKEEDQFYVMQVICOQLIYKTIQSLPKA 266
 Db 1192 SI-----GIMLEELARIYKGEQLPDLRLQYKDYAVWQ--SRQAEGYKKDQAYWKE 1240
 QY 267 L--QERYLLVRVEDLARAPV 284
 Db 1241 VFAGELPVLQLSDYPRPV 1260
 RESULT 5
 TP5B_CAEEL STANDARD; PRT; 359 AA.
 ID TP5B_CAEEL

RESULT 5

MACF_HUMAN	STANDARD;	PRT;	5430 AA.
Q9P9N3;	Q9WXY2;	Q9H540;	Q9ULG9;
16-OCT-2001	(Rel. 40, Created)		
16-OCT-2001	(Rel. 40, Last sequence update)		
16-OCT-2001	(Rel. 42, Last annotation update)		
Microtubule-actin crosslinking factor 1, isoforms 1/2/3 (Actin cross-			
linking family protein 7) (Macrophin 1) (Trabeculin-alpha) (620 kDa			
actin-binding protein) (ABP620)			
MACF1 OR AC7 OR ABP620 OR KIA0465 OR KIAA1251.			
Hom sapiens (Human)			
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI TaxID=9606;			
[1]			
SEQUENCE FROM N.A. (ISOFORM 2).			
MEDLINE=20001959; PubMed=10529403;			
Okada T., Matsuda S., Nakatagawa S., Ichigotani Y., Iwahashi N.,			
Takahashi M., Ishigaki T., Hamaguchi M.;			
"Molecular cloning of macrophin, a human homologue of Drosophila			
kapapo with a close structural similarity to plectin and dystrophin."			
Biochem. Biophys. Res. Commun. 264:568-574(1999).			
[2]			
SEQUENCE FROM N.A. (ISOFORM 1).			
MEDLINE=20026884; PubMed=10559237;			
Sun Y., Zhang J., Krafetz S.-K., Auchair D., Chang M.-S., Liu Y.,			
Sutherland R., Saigia R., Griffin J.D., Ferland L.H., Chen L.B.;			
"Molecular cloning and characterization of human trabeculin-alpha, a			
giant protein defining a new family of actin-binding proteins."			
J. Biol. Chem. 274:33522-33530(1999).			
[3]			
SEQUENCE OF 182-4812 FROM N.A. AND ALTERNATIVE SPLICING.			
MEDLINE=21833812; PubMed=11845288;			
Gong T.-W.L., Besirli C.G., Lomax M.I.;			
"MACF1 gene structure: a hybrid of plectin and dystrophin."			
Mamm. Genome 12:852-861(2001).			
[4]			
SEQUENCE OF 868-2350 FROM N.A.			
TISSUE=Brain;			
MEDLINE=20039619; PubMed=10574462;			
Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,			
Ohara O.;			
"Prediction of the coding sequences of unidentified human genes. XV.			
The complete sequences of 100 new cDNA clones from brain which code			
for large proteins in vitro."			
DNA Res. 6:337-345(1999).			
[5]			
SEQUENCE OF 1544-5057 FROM N.A.			
Corby N.;			
Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
[6]			
SEQUENCE OF 3312-5430 FROM N.A.			
TISSUE=Brain;			
MEDLINE=212158633; PubMed=12168954;			
Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;			
"Construction of expression-ready cDNA clones for KIAA genes: manual			
curation of 330 KIAA cDNA clones."			
DNA Res. 9:99-106(2002).			
[7]			
SEQUENCE OF 3734-5430 FROM N.A.			
TISSUE=Brain;			
MEDLINE=98116662; PubMed=9455484;			
Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,			
Nakajima D., Nomura N., Ohara O.;			
"Characterization of cDNA clones in size-fractionated cDNA libraries			
from human brain."			
DNA Res. 4:1345-349(1997).			
-1- LINKING: F-actin-binding protein which may play a role in cross			
linking reaction to other cytoskeletal proteins. Also binds to			
microtubules (By similarity).			
-1- SUBCELLULAR LOCATION: Cytoplasmic.			

-1- ALTERNATIVE PRODUCTS.

CC Event=Alternative splicing; Named isoforms=4;

CC Name=2;

CC IsoId=Q9UPN3-2; Sequence=Displayed;

CC Name=1;

CC IsoId=Q9UPN3-1; Sequence=VSP_007341;

CC Name=3;

CC IsoId=Q9UPN3-3; Sequence=Not described;

CC Name=4;

CC IsoId=Q9EPK2-1; Sequence=External;

CC -1- TISSUE SPECIFICITY: Ubiquitously expressed

CC -1- SIMILARITY: Belongs to the plakin or cytolinker family.

CC -1- SIMILARITY: Contains 1 actin-binding domain.

CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.

CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.

CC -1- SIMILARITY: Contains 1 SH3 domain.

CC -1- SIMILARITY: Contains 37 spectrin repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB029290; BAA83821.1; -

DR EMBL; AF141968; AAF06360.1; -

CC EMBL; AF325341; AAL39000.1; -

DR EMBL; AF325341; AAL39000.1; JOINED.

CC EMBL; AF325330; AAL39000.1; JOINED.

DR EMBL; AF325331; AAL39000.1; JOINED.

CC EMBL; AF325332; AAL39000.1; JOINED.

DR EMBL; AF325333; AAL39000.1; JOINED.

CC EMBL; AF325334; AAL39000.1; JOINED.

DR EMBL; AF325335; AAL39000.1; JOINED.

CC EMBL; AF325336; AAL39000.1; JOINED.

DR EMBL; AF325339; AAL39000.1; JOINED.

CC EMBL; AF325340; AAL39000.1; JOINED.

DR EMBL; AB033077; BAA86565.1; -

CC EMBL; ALJ37853; CAC15920.1; -

DR EMBL; AB007934; BAA32310.2; -

CC PIR; T00079; T00079.

DR HSSP; Q01082; 1BKR.

CC Genew; HGNC:13684; MACF1.

DR GO; GO:0005856; Cytoskeleton; NAS.

CC GO; GO:0003780; F:actin cross-linking activity; NAS.

DR GO; GO:0005509; F:calcium ion binding; NAS.

CC GO; GO:0008017; F:microtubule binding; NAS.

DR InterPro; IPR001589; Actbind actinin.

CC InterPro; IPR001715; Calponin-like.

DR InterPro; IPR002048; EF-hand.

CC InterPro; IPR003108; GAS2.

DR InterPro; IPR004452; SH3.

CC InterPro; IPR002017; Spectrin.

DR Pfam; PF00307; CH; 2.

CC Pfam; PF00036; ehand; 2.

DR Pfam; PF02187; GAS2; 1.

CC Pfam; PF00435; spectrin; 27.

DR ProDom; PD000012; EF-hand; 1.

CC SMART; SM00033; CH; 2.

DR SMART; SM00054; Eph; 2.

CC SMART; SM00243; GAS2; 1.

DR SMART; SM00150; SPEC; 36.

CC PROSITE; PS00019; ACTIN1; 1.

DR PROSITE; PS00020; ACTIN2; FALSE_NEG.

CC PROSITE; PS00021; CH; 2.

DR PROSITE; PS00018; CH; 2.

CC PROSITE; PS00002; SH3; FALSE_NEG.

DR Actin-binding; Cytoskeleton; Calcium; Calcium-binding; Repeat;

KW SH3 domain; Coiled coil; Alternative splicing.

FT DOMAIN 1 295

FT FT 78 181

FT DOMAIN 194 295

FT CH 1.

FT CH 2.

FT

QY 174 -VVLKEVFFNLOSLLYPLLLKOPSLNLHIVLRDPRVFRSRETKGLMIDSR----- 226
 Db 1874 FWTISGQVLDMEFKEGKEPS-----EIGNLVKD-----KUKDATERVTAHSHKCTRIGS 1925
 QY 227 ---IYMGHEQKLKEDQFYVMQVICSQLE--IYKTIQSLPKALQERYLLVRYEDLAR 281
 Db 1926 HLNMLLGQVHQFQNSADSLQAMMQA-CEANVEKLLSDTVASDPGVLSQ----- 1973
 QY 282 APVAOTSMYEVVGLFEP--HLQTVHNI--TRQKMGDHFHTNAEDA-LNVYQAWRW 336
 Db 1974 --LATTQQLQELAEHQVPEKLVQKQVARDIMEIEGEPADPHRHVQVETDTSILSHFQSLSY 2031
 QY 337 SLPEYKVKSRQLQACDAMNLLGYRVRSEQRNLLL 373
 Db 2032 SLA-ERSILOXAIQAQ-----QSVQESLESLL 2058
 RESULT 7
 MAC4 HUMAN
 ID MAC4 HUMAN STANDARD; PRT; 5938 AA.
 AC Q96PK2; Q8WKY1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Microtubule-actin crosslinking factor 1, isoform 4.
 GN MACF1 OR ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Buthera; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RX MEDLINE=21833812; PubMed=11845288;
 RA Gong T.-W.L., Besirli C.G., Lomax M.I.;
 RT "MACF1 gene structure: a hybrid of plectin and dystrophin.";
 RL Mamm. Genome 12:852-861(2001).
 CC -!- FUNCTION: May play a role in cross-linking cytoskeletal proteins
 CC by binding intermediate filaments to the N-terminal plectin
 CC repeats and microtubules to the C-terminus.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=4;
 CC IsoId=Q96PK2-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q9UPN3-1; Sequence=External;
 CC Name=2;
 CC IsoId=Q9UPN3-2; Sequence=External;
 CC Name=3;
 CC IsoId=Q9UPN3-3; Sequence=External;
 CC -!- TISSUE SPECIFICITY: Expressed in heart, lung, pituitary and
 CC placenta, not found in brain, kidney, liver, pancreas or skeletal
 CC muscle.
 CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -!- SIMILARITY: Contains 19 plectrin repeats.
 CC -!- SIMILARITY: Contains 32 spectrin repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF317696; AAL09459.1; -;
 DR EMBL; AF325341; AAL38997.1; -;
 DR EMBL; AF325333; AAL38997.1; JOINED.
 DR EMBL; AF325334; AAL38997.1; JOINED.
 DR EMBL; AF325335; AAL38997.1; JOINED.
 DR EMBL; AF325336; AAL38997.1; JOINED.

DR EMBL; AF325339; AAL38997.1; JOINED.
 DR EMBL; AF325340; AAL38997.1; JOINED.
 DR GO; GO:0005856; Cytoskeleton; ISS.
 DR GO; GO:0005809; F-actin ion binding; ISS.
 DR GO; GO:0008017; F-actin tubule binding; ISS.
 DR InterPro; IPR003108; GAS2.
 DR InterPro; IPR001101; Plectrin repeat.
 DR InterPro; IPR02017; Spectrin.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF02187; GAS2; 1.
 DR Pfam; PF00681; Plectrin; 11.
 DR Pfam; PF00435; spectrin; 26.
 DR ProDom; PD000012; EF-hand; 1.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00243; GAS2; 1.
 DR SMART; SM00250; PLEC; 19.
 DR SMART; SM00150; SPEC; 32.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR Cytokeleton; Calcium; Calcium-binding; Repeat; Coiled coil;
 KW Alternative splicing.
 FT DOMAIN 1830 1936 COILED COIL (POTENTIAL).
 FT DOMAIN 2001 2192 COILED COIL (POTENTIAL).
 FT DOMAIN 2282 2345 COILED COIL (POTENTIAL).
 FT DOMAIN 2477 2507 COILED COIL (POTENTIAL).
 FT DOMAIN 2541 2654 COILED COIL (POTENTIAL).
 FT DOMAIN 2686 2814 COILED COIL (POTENTIAL).
 FT DOMAIN 2887 2919 COILED COIL (POTENTIAL).
 FT DOMAIN 3046 3197 COILED COIL (POTENTIAL).
 FT DOMAIN 3262 3503 COILED COIL (POTENTIAL).
 FT DOMAIN 3632 3666 COILED COIL (POTENTIAL).
 FT DOMAIN 3746 3779 COILED COIL (POTENTIAL).
 FT DOMAIN 3920 3984 COILED COIL (POTENTIAL).
 FT DOMAIN 4098 4168 COILED COIL (POTENTIAL).
 FT DOMAIN 4288 4308 COILED COIL (POTENTIAL).
 FT DOMAIN 4354 4386 COILED COIL (POTENTIAL).
 FT DOMAIN 4397 4433 COILED COIL (POTENTIAL).
 FT DOMAIN 4469 4489 COILED COIL (POTENTIAL).
 FT DOMAIN 4586 4720 COILED COIL (POTENTIAL).
 FT DOMAIN 4845 4880 COILED COIL (POTENTIAL).
 FT DOMAIN 4910 4939 COILED COIL (POTENTIAL).
 FT DOMAIN 4970 5000 COILED COIL (POTENTIAL).
 FT DOMAIN 5409 5437 COILED COIL (POTENTIAL).
 FT DOMAIN 5546 5569 COILED COIL (POTENTIAL).
 FT CA_BIND 5598 5610 EF-HAND 1 (POTENTIAL).
 FT CA_BIND 5634 5646 EF-HAND 2 (POTENTIAL).
 FT REPEAT 12 49 PLECTIN 1.
 FT REPEAT 53 88 PLECTIN 2.
 FT REPEAT 89 126 PLECTIN 3.
 FT REPEAT 130 164 PLECTIN 4.
 FT REPEAT 166 202 PLECTIN 5.
 FT REPEAT 203 240 PLECTIN 6.
 FT REPEAT 243 278 PLECTIN 7.
 FT REPEAT 279 316 PLECTIN 8.
 FT REPEAT 318 354 PLECTIN 9.
 FT REPEAT 725 762 PLECTIN 10.
 FT REPEAT 763 800 PLECTIN 11.
 FT REPEAT 801 838 PLECTIN 12.
 FT REPEAT 839 876 PLECTIN 13.
 FT REPEAT 897 934 PLECTIN 14.
 FT REPEAT 935 972 PLECTIN 15.
 FT REPEAT 975 1011 PLECTIN 16.
 FT REPEAT 1012 1049 PLECTIN 17.
 FT REPEAT 1121 1158 PLECTIN 18.
 FT REPEAT 1159 1196 PLECTIN 19.
 FT REPEAT 1925 2032 SPECTRIN 1.
 FT REPEAT 2052 2160 SPECTRIN 2.
 FT REPEAT 2211 2313 SPECTRIN 3.
 FT REPEAT 2320 2430 SPECTRIN 4.
 FT REPEAT 2437 2543 SPECTRIN 5.
 FT REPEAT 2550 2652 SPECTRIN 6.
 FT REPEAT 2659 2758 SPECTRIN 7.
 FT REPEAT 2765 2896 SPECTRIN 8.

FT REPEAT 2903 3008 SPECTRIN 9.
FT REPEAT 3015 3119 SPECTRIN 10.
FT REPEAT 3126 3229 SPECTRIN 11.
FT REPEAT 3236 3339 SPECTRIN 12.
FT REPEAT 3346 3446 SPECTRIN 13.
FT REPEAT 3453 3555 SPECTRIN 14.
FT REPEAT 3562 3664 SPECTRIN 15.
FT REPEAT 3671 3775 SPECTRIN 16.
FT REPEAT 3782 3884 SPECTRIN 17.
FT REPEAT 3891 3993 SPECTRIN 18.
FT REPEAT 4000 4102 SPECTRIN 19.
FT REPEAT 4109 4211 SPECTRIN 20.
FT REPEAT 4218 4320 SPECTRIN 21.
FT REPEAT 4327 4428 SPECTRIN 22.
FT REPEAT 4438 4544 SPECTRIN 23.
FT REPEAT 4551 4653 SPECTRIN 24.
FT REPEAT 4660 4763 SPECTRIN 25.
FT REPEAT 4770 4872 SPECTRIN 26.
FT REPEAT 4879 4982 SPECTRIN 27.
FT REPEAT 4989 5091 SPECTRIN 28.
FT REPEAT 5098 5201 SPECTRIN 29.
FT REPEAT 5208 5309 SPECTRIN 30.
FT REPEAT 5316 5418 SPECTRIN 31.
FT REPEAT 5425 5555 SPECTRIN 32.
FT CONFLICT 1712 1712 P -> S (IN REF. 1; AAL38997).
SQ SEQUENCE 5938 AA; 670134 MW; B8784112752DA004 CRC64;

Query Match 4.5%; Score 91.5; DB 1; Length 5938;
Best Local Similarity 22.2%; Pred. No. 87;
Matches 88; Conservative 66; Mismatches 128; Indels 115; Gaps 24;

QY 2 2 LIPKMKLLFLVSOAILALFFHYSHNISLSMKAPERMHVILVSSWRSGSFVQQL 61
Db 2254 LLSQQQNFIATQSAQA-----FLDQGHNLT-----PEEQML-----QQKLGL 2294

QY 62 FQCHDFVFLMEPAHVMTFQSTAWMLHM-AVDLIRAVFLCDMSVDFVMEPGPRQ 120
Db 2295 KEQYS-----TSLAQSAELKQVOTLQELQK-FLQHKKEFSWLE---RSE 2337

QY 121 SLSFOWENSRLACSPACDIIPQDEIIPRAHRLCSQPPF--EYVEKACRSYSH- 173
Db 2338 KEL---ENMHKGGSP-----ETLP-----SLLRKQGSFSDVI-----SHKGLR 2375

QY 174 -VVLKEVRFNLSLYPLKDSMLNHVILVDRPRAVRSRERKGLMIDSR----- 226
Db 2376 FVTISQKVLDMENSGFKGKPS-----EIGNLYKD-----KLKDATERVTLHSCITRLGS 2427

QY 227 ---IVMGQHEQKLKEDQPYVWQVVICOSQLF--IYKTIQSLPKALQERYLLVRYEDLAR 281
Db 2428 HLNMLLGQYHQFQNSADSLOAMQA-CEANVEKLSDIVASDPGVLOEQ----- 2475

QY 282 AVPAQTSRMYEFGLEFLP--HLQTVWNI--TRKGMGDHAFTNARDA-LNVSQAWRW 336
Db 2476 --LATTQQLQELAEHQVVEKLVQKQVARDIMEIEGEPAPDRHRVQETTDLSLHFQSLSY 2533

QY 337 SLPEYKVSRLQACGADNALLGLVHRSRQEQRNLL 373
Db 2534 SLA-ERSLLQKRAIQS-----QSQESLESLL 2560

RESULT 8
ID -239F HUMAN STANDARD; PRT; 294 AA.
AC Q1577;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Fetal brain protein 239 (239F8).
GN C11ORF8 OR FAM1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

RN [1] SEQUENCE FROM N.A.
RX MEDLINE=95080775; PubMed=7527372;
RA Schwartz F., Neve R., Eisenman R., Gessler M., Bruns G.;
RT "A WAGR region gene between PAX-6 and FSHB expressed in fetal brain.";
RL Hum. Genet. 94:658-664(1994).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=96115606; PubMed=8666403;
RA Schwartz F., Eisenman R., Knoll J., Gessler M., Bruns G.;
RT "cDNA sequence, genomic organization, and evolutionary conservation
of a novel gene from the WAGR region.";
RL Genomics 29:526-532(1995).
RN [3]
RX SEQUENCE FROM N.A.
RC TISSUE=Colon, Kidney, and Stomach;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN FETAL BRAIN.
CC -!- SIMILARITY: BELONGS TO THE UPF0046 FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U57911; AAC50564.1; --
DR EMBL; BC031582; AAH31582.1; --
DR Genbank; HGNC:1180; C11orf8.
DR MIM; 600911; --
DR GO; GO:0007399; P:neurogenesis; TAS.
DR InterPro; IPR004843; M-ppestrase.
DR Pfam; PF00149; Metallophos; 1.
SQ SEQUENCE 294 AA; 33360 MW; 43B2BC0DA1BFD1F0 CRC64;

Query Match 4.5%; Score 91; DB 1; Length 294;
Best Local Similarity 20.2%; Pred. No. 2.3;
Matches 53; Conservative 30; Mismatches 89; Indels 90; Gaps 9;

QY 139 DIIPQDEIIPRAHRLCSQPPFVEVKEACRSYSHVL-----KEVRFNLSQL 187
Db 45 DIPYDTPKPAQTRFVCSIDTSHRTDGIQMPYGDILLHTGDTGLPSEVKFN----- 100

QY 188 YPLLKDFSLNHVILVDRPRAVRSRERKGLMIDSRVIM-GQHEQKLKKE-----D 240
Db 101 -----DWLGNLPYEYKIVLVIAGNHELFDKEFMADLVK 132

QY 241 QPYVWQVVICOSQLEIYKTIQSLPKALQERYLLVRYEDLARPAVQTSRM-----YEFVGL 296
Db 133 QDYVRRPSPVSKLPEDPDNVQSL-----LTNSIYLODSEVTVKGFRIYGA 177

```

QY 297 EFLPHLQVHNIITGCKMGDHPHTNARDALNVSOAWSLPYEKYSRLOKACGDAMNL 356
DB 178 PWTWFWNGWGNLPRQGLD-----KWNLIPEGIDILM--THGPP--- 216
QY 357 LGYHRVSEGEQQRNLLDLST 378
DB 217 LGFRDWVPELQVGCVELLNT 238

RESULT 9
MX3_RAT
ID MX3_RAT STANDARD; PRT; 659 AA.
AC P18590;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Interferon-induced GTP-binding protein MX3.
GN MX3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91056598; PubMed=2173790;
RA Meier E., Kunz G., Haller O., Arnheiter H.;
RT "Activity of rat Mx proteins against a rhabdovirus.";
RL J. Virol. 64:6263-6269(1990).
CC -!- FUNCTION: Does not show activity against influenza virus or VSV;
CC although it only differs from Mx2 by 8 positions.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- INDUCTION: By interferons.
CC -!- SIMILARITY: Belongs to the dynamin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52713; CAA36937.1; -
CC PIR; S11737; S11737.
CC InterPro; IPR001401; Dynamin.
CC InterPro; IPR000375; Dynamin_central.
CC InterPro; IPR003130; GED.
CC Pfam; PF00350; dynamin; 1.
CC Pfam; PF01031; dynamin_2; 1.
CC Pfam; PF02212; GED; 1.
CC PRINTS; PR00195; DYNAMIN.
CC SMART; SM00053; DINC; 1.
CC SMART; SM00302; GED; 1.
CC Interferon induction; GTP-binding; Multigene family.
CC PROSITE; PS00410; DYNAMIN; 1.
CC NP_BIND 75 82 GTP (POTENTIAL).
CC NP_BIND 176 180 GTP (POTENTIAL).
CC NP_BIND 245 248 GTP (POTENTIAL).
CC SEQUENCE 659 AA; 74951 MW; 755D63283BC865B3 CRC64;

Query Match 4.5%; Score 91; DB 1; Length 659;
Best Local Similarity 20.6%; Pred. No. 6.2;
Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;

QY 62 FGOHPDVFLMEPAWVWTFKQSTAWMLHMAVRDLI-----RAVFLCDMSVFDAYME--- 114
DB 301 FKEHPQFALLEDG-----RATVECLAERTMTLISHICKSLPLENQIKESHQSYSE 353
QY 115 -----PGPRQSSLFOWNSRALCSAPACDIIPODEIIPRAHCRLLCS-QQPF--- 161
DB 354 ELQKYGADIPEDENEXTLFLEKINAF-NQDITAVEGEIYREKECRFLTKLRKEFFLW 412
QY 162 -EVVEKACRSYSHVIVKEVRFNLFQ-----SLYPLLKDPSLNLIH 200

```

```

DB 413 SEIERNFQKGDALYKEVYIFEMQYRGRELPGFVNYKTFENIRROIKTLEEPAMEM-- 470
QY 201 VHLVRD-PAVPR-----SRRTKGLDMLTDSIRVMGQHEQKLKEDQPYVMQV 248
DB 471 LHKVTEIVRAAFTTVSEKNFSEFFNLHRTTKSKEDIRL---EQTEAEKAILRHQFMEQ 527
QY 249 ICQSLEIYKTIQSLPKALQERYLLVRYEDL-----ARAPVAQTSRMVYFVGL 296
DB 528 IYQCQDIYR-----KALQK-----VREEEAEEERKHKGRSQAQSPNLQTSMD----- 572
QY 297 EFLPHLQVHNIITGCKMGDHPHTNARDALNVSOAWSLPYEKYSRLOKACGDAMNL 356
DB 573 EIPQHLNAYRQBAHN 587

RESULT 10
C7D8_SOYBN
ID C7D8_SOYBN STANDARD; PRT; 504 AA.
AC O81974;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome P450 71D8 (EC 1.14.--) (P450 CP7).
GN CYP71D8.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eusids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Harosoy 63;
RX MEDLINE=98311068; PubMed=9648734;
RA Schopfer C.R., Ebel J.;
RT "Identification of elicitor-induced cytochrome P450s of soybean
RT (Glycine max L.) using differential display of mRNA.";
RL Mol. Gen. Genet. 258:315-322(1998).
CC -!- INDUCTION: By fungal elicitor.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y10493; CAA71517.1; -
CC PIR; T07120; T07120.
CC HSP; P14779; IJPZ.
CC InterPro; IPR001128; Cytochrome_P450.
CC Pfam; PF00067; P450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Heme.
CC METAL 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SEQUENCE 504 AA; 57579 MW; 86E4FAEBC2FF2A8F CRC64;

Query Match 4.4%; Score 89.5; DB 1; Length 504;
Best Local Similarity 19.9%; Pred. No. 5.9;
Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16;

QY 68 VEVLMPEAWVWVWTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRQSSLFOWE 127
DB 14 VFLDLH--WLV-KTYKQKSHKL-----PPGPR----- 39
QY 128 NSRALSAPACDIIPODEIIPRAHCRLLCSQQPFVVEKACRSYS---HVVLEKVEFFNL 184
DB 40 -----LP-----IIGNLHQLAALASLPQALQKLVRYKGYPLMLHQLGEISTLV 83
QY 185 QSLYPLLKDPSLNLIHIVHLVRDPAVPRSRRTKGLMIDSRIVMGQHEQKLKEDQPY 244

```

```

Db 84 SS--PKAMENMKTHDVHVRPQ-----LLAPQFMYGATDIAFAPYGDYR 129
Qy 245 VMQVCSQSLIYKTIQSLPALQE--RYLLRVVEDIARAPVACTSMRYEVLG----- 296
Db 130 QIRKICTLELSAKRVOSFHSIRODENKKLQTSIHSSAGSPIDLSGKFLSLGTTVSRAA 189
Qy 297 -----EFL-----PHLOTWVHNITRKGMDGDAFHNTA----- 324
Db 190 FGKENDODEFMSLVRKAITWGGFEVDMDPEPSLKP-LHLLTROKAKVEHV-HQRADKIL 247
Qy 325 RDALNVSOQWMSLPYKRVSLQKACGAMNLLGYRHRVSEFQBNLLDL 376
Db 248 EDILRKHM-----EKRTVKSGNG-----SEAEQEDLVLL 279

```

RESULT 11

```

PKSM_BACSU
ID PKSM_BACSU STANDARD; PRT; 4273 AA.
AC P40872; O31781;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative polyketide synthase pksm.
CN PKSM OR PKSY OR BSU7200.
OS Bacillus subtilis.
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
EX MEDLINE=9804033; PubMed=9384377;
RA Kunst E., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.P., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Britington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeda T., Takemaru K., Tanaka T., Terpatra P., Tognoni A.,
RA Tostato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
RN [2]
RP SEQUENCE OF 1-1763 FROM N.A.
RC STRAIN=168 / PB1424;
RA Tognoni A., Grandi G.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM.
CC -!- COFACTOR: Contains 4 covalently bound phosphopantetheines
CC (Potential).

```

```

CC -!- SIMILARITY: Contains 4 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z99113; CAB13603.1; -
CC EMBL; Z35133; CA84505.1; -
CC PIR; C69679; C69679.
CC Subtilisin, BG10931; pksm.
CC InterPro; IPR002198; ADH short.
CC InterPro; IPR000794; Ketoacyl synth.
CC InterPro; IPR001601; Methyltransf.
CC InterPro; IPR006163; Pp bind.
CC InterPro; IPR006162; Ppantne S.
CC InterPro; IPR000051; SAM_bind.
CC Pfam; PF00106; adh_short; 1.
CC Pfam; PF00109; ketoacyl-synt; 3.
CC Pfam; PF02801; ketoacyl-synt C; 3.
CC Pfam; PF00550; pp-binding; 4.
CC PROSITE; PS00012; PHOSPHOPANTHEINE; 2.
CC PROSITE; PS00806; B_KETOACYL SYNTHASE; 2.
CC PROSITE; PS00075; ACP_DOMAIN; 4.
KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 295 364
FT ACYL CARRIER (ACP) 1.
FT DOMAIN 396 834
FT ACYL CARRIER (ACP) 2.
FT DOMAIN 2190 2258
FT BETA-KETOACYL SYNTHASE 2.
FT DOMAIN 2322 2737
FT BETA-KETOACYL SYNTHASE 3.
FT DOMAIN 3532 3947
FT ACYL CARRIER (ACP) 3.
FT DOMAIN 3410 3483
FT ACYL CARRIER (ACP) 4.
FT DOMAIN 4140 4203
FT PHOSPHOPANTHEINE (POTENTIAL).
FT BINDING 327 327
FT BINDING 2222 2222
FT ACT SITE 2476 2476
FT BINDING 3446 3446
FT ACT SITE 3690 3690
FT BINDING 4172 4172
FT CONFLICT 103 103
FT CONFLICT 276 276
FT CONFLICT 289 289
FT CONFLICT 289 289
SQ SEQUENCE 4273 AA; 477459 MW; 3BBFCF1A250AEB5A CRC64;
Query Match 4.4%; Score 89.5; DB 1; Length 4273;
Best Local Similarity 22.9%; Pred. No. 85;
Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;
Qy 161 FEVVEKACRSYSHVWLKEVFFENLQSLYPLKDPD-----LNLHIVHLVRDPAVRSRER 216
Db 36 YQVFQEHGAYQLELEK-----NLTFYPLADESYDIALTIHVSEEGTWSIIIDGQK 90
Qy 217 TKGDLMDISRIYVNGQHE-QKLKKEQPPYVMQVICOQSLKYKTIQSLPALQERYLLVR 275
Db 91 QHGESLSDXR-----QYETADMRKQETAFAESI-----DLNOWNKSTADRLNLDIYEOCR 142
Qy 276 YEDLARAPVAQT-SRMYE-----FVGLFELPHLOTWVHNIT--RGKGMGDHAFHTN 323
Db 143 SQELVHTGMKREGQIYEKEGAVIDLAVGQALRHSDAFLPHPLIDSGSIGG----- 196
Qy 324 ARDALNVSOQWMSLPYKRVSLQKAC 350
Db 197 --SCLTSDQTMVLPVLYESFSASERLQKGC 224

```

RESULT 12

```

P4ST FLACH
ID P4ST FLACH
AC P52837;
DT 01-OCT-1996 (Rel. 34, Created)

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Flavonol 4'-sulfotransferase (EC 2.8.2.-) (F4-ST).
 OS Flaveria chloraefolia.
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Asterales; Asteraceae; Asteroideae; Tageteae; Flaveria.
 OX NCBI_TaxID=4228;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92159034; PubMed=1741382;
 RA Varin L., Deluca V., Ibrahim R.K., Brisson N.;
 RT "Molecular characterization of two plant flavonol sulfotransferases.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).
 RN [2].
 RP PAPS-BINDING SITE.
 RX MEDLINE=95279378; PubMed=7759495;
 RA Varin L., Marsolais F., Brisson N.;
 RT "Chimeric flavonol sulfotransferases define a domain responsible for
 RT substrate and position specificities.";
 RL J. Biol. Chem. 270:12498-12502(1995).
 CC -!- FUNCTION: Transfers sulfate group into flavonol position 4'.
 CC May play a role in auxin transport.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- TISSUE SPECIFICITY: Highest in shoot tips and lowest in mature
 CC leaves and roots (By similarity).
 CC -!- SIMILARITY: Belongs to the plant sulfotransferase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC EMBL; M841136; AAA33343.1; --
 DR HSSP; P50224; ICMJ.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR ProDom; PD001218; Sulfotransferase; 1.
 KW Transferase.
 FT BINDING 148 163 PAPS-BINDING SITE.
 SQ SEQUENCE 320 AA; 37255 MW; 5B263659F6CCBCC0 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 320;
 Best Local Similarity 20.8%; Pred. No. 3.7;
 Matches 56; Conservative 30; Mismatches 83; Indels 100; Gaps 12;
 QY 54 GSFVQLFGQHP-DVEYLMPEAWHVMFTFKOSTAWMLHMAVRLIRAVFLCDMSVFDAY 112
 DB 47 GAILAQSFKARDDDVFLCSYP-----KSGTTLKALAVIVR-----EKDFEF 91
 QY 113 MEP-----GPRQSSLFQ-WENSRALCAPACDIIIPQDEIIPRAHCRLLC-----SQPFVEVE 165
 DB 92 TSPLLATNPICPIPIYEDKDKTKIVENQNSCTP-----MATHMPYH 133
 QY 163 VVEKACRSYSHVVLKEVRFNLSQLYPLKDPNLNLIHVLRDPAVRSRERTKGDLM 222
 DB 134 VLPKS-----IL-----ALNCKMVIYTRINDV----- 156
 QY 223 IDSRIYMGQHEQKLKEDQPYVMQVICOSEIYKTIQSLF-----KALQER-- 270
 DB 157 IVSFYHFGREITKLPEDAPFE-----EAFDEFYHGISQFGFYWDHLLGYKASLERPE 210
 QY 271 -YLLVYEDLARAPVACTSRMYEFCVLEF 298
 DB 211 VILFLKYEDVKDPTSNVRLAEFIGYFF 239
 RESULT 13
 ID MPPA_RAT STANDARD; PRT; 524 AA.

P20069;
 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mitochondrial processing peptidase alpha subunit, mitochondrial
 DE precursor (EC 3.4.24.64) (Alpha-MPP) (P-55).
 GN MPPCA OR INP5E OR MPPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1].
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=91045920; PubMed=2236012;
 RA Kleiber J., Kalousek F., Swaroop M., Rosenberg L.E.;
 RT "The general mitochondrial matrix processing protease from rat liver:
 RT structural characterization of the catalytic subunit.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7978-7982(1990).
 CC -!- FUNCTION: The mitochondrial processing protease (MPP-I) cleaves
 CC presequences from mitochondrial protein precursors. Most MPP-I
 CC cleavage sites follow an arginine at position -2.
 CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
 CC precursor proteins imported into the mitochondrion, typically with
 CC Arg in position P2.
 CC -!- SUBUNIT: Heterodimer of alpha and beta subunits.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- SIMILARITY: Belongs to peptidase family M16.
 CC -!- CAUTION: Does not seem to have a protease activity as it lacks the
 CC zinc-binding site.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC EMBL; M57728; AAA41632.1; --
 DR PIR; A36205; A36205.
 DR MEROPS; M16.971; --
 DR InterPro; IPR001431; Peptidase M16.
 DR InterPro; IPR007863; Peptidase_M16_C.
 DR Pfam; PF00675; Peptidase_M16; 1.
 DR Pfam; PF05193; Peptidase_M16_C; 1.
 DR PROSITE; PS00143; INSULINASE; 1.
 KW Hydroxylase; Metalloprotease; Mitochondrion; Transit peptide.
 FT TRANSIT 1 32 MITOCHONDRION.
 FT CHAIN 33 524 MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA
 FT SUBUNIT
 SQ SEQUENCE 524 AA; 58607 MW; 8BF08FBC9FF09DB2 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 524;
 Best Local Similarity 21.5%; Pred. No. 6.8;
 Matches 70; Conservative 46; Mismatches 104; Indels 106; Gaps 19;
 QY 70 YLM--EPAW---HYVMTFKQ-----STAWMLHMAVRLIRAVFLCDMSVFDAYMEPG 116
 DB 269 YLGVQPAWGAQAVWMLTAQWHSTGGSRW-----RE-----TCQMSA---LRP- 311
 QY 117 PRQSSLFQWENSRALCAPACDIIIPQDEIIPRAHCRLLC-----SQPFVEVE 165
 DB 312 PRQSSHI-YGGAREL-----LLEEDFIPPAVLNMMWGGGSGSAGPGKGMFS--- 360
 QY 166 KACRSYSHVVLKEVRFNLSQLYPLKDPNLNLIHVLRDPAVRSRERTKGDLMIDS 225
 DB 361 ---RLYLNVLNRHHWMYNATSYHSHVED--TGLLCITHASADPROV-----REMVETITKE 410
 QY 226 RIVMGQHEQKLKEDQPYVMQVICOSEIYKTIQSLPKALQERVLLVRYEDLARAPVA 285
 DB 411 FILMGRTVDLVEIERAKTQLMSLMN-----LESRP-----VIFEDVGQVLA 454

```
QY 286 QTSRMVEYFGLFLPH-LQTVHVNITRGKMGDHPHTNARDALNVSQAWWSLPIYEKVS 344
Db 455 TSHRK-----LPHELCTLRNV-----KPEDIKRVASKMLRGKP----- 488
QY 345 RLQKACGDAMNLLGYRHVRSQEQRN 370
Db 489 -AVALGDLTDLPTTEHIOALSND 513

RESULT 14
CAPT STRPY STANDARD; PRT; 920 AA.
ID Q9A0U7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPcase) (PEPC).
GN PPC OR SPY0608.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Aidic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Clifton F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Giffon S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -!- FUNCTION: Through the carboxylation of phosphoenolpyruvate (PEP)
CC it forms oxaloacetate, a four-carbon dicarboxylic acid source for
CC the tricarboxylic acid cycle.
CC -!- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +
CC phosphoenolpyruvate + CO(2).
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SIMILARITY: Belongs to the PEPcase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006516; AAK33584.1; -
CC HSSP; P00884; 1FY. -
CC HAMAP; MF00595; -; 1.
CC InterPro; IPR001449; PEPcase.
CC Pfam; PF00311; PEPcase; 1.
CC PRINTS; PR00150; PEPCARBXLASE.
CC PROSITE; PS00781; PEPcase 1; 1.
CC PROSITE; PS00393; PEPcase 2; 1.
CC Lyase; Carbon dioxide fixation; Tricarboxylic acid cycle;
CC Complete proteome.
CC ACT_SITE 138 138 BY SIMILARITY.
CC ACT_SITE 583 583 BY SIMILARITY.
CC SEQUENCE 920 AA; 104751 MW; 464EA4A309A22237 CRC64;

Query Match 4.4%; Score 89; DB 1; Length 920;
Best Local Similarity 18.6%; Pred. No. 14;
Matches 49; Conservative 48; Mismatches 85; Indels 82; Gaps 15;

QY 157 SQPF-----EVEKACRSYSHVVLKEVRFNLOSFLYLLKDPSLNLIHVL 203
Db 634 TSQPFSGIKDRILRTQGBIIE---NKYGN---KDVAYVHLE---MLISASINRMVQM 683
QY 204 VRDPAVFSRERTKGLMIDSIYVQGEOKKKEDOPYVYVQVICOQLEIYKTIQSL 263
Db 684 ITDPEIDSFRE-IMDSIVADSNIIY-----RKL-VFDNPHFYDYFEASPI-----KEVSSL 734
```

```
QY 264 PRALQERYLLVREYEDLAPARVAQTSRMVEYFGLFLPHLQTVHVN----- 308
Db 735 -----NIGSPAARKT-ITETGLRAIPWVFSWQNRIMFPGWYGVGSAPK 779
QY 309 ---ITRGKMGD---HAFHTNARDALNVSQAWWSLPIYEKVSRLQKACGDA-MNL-LGYRH 361
Db 780 RYIDRAQGNLERLQHYOT-----WPFHSLLSNVDMVLSKSNNTAFQY 826
QY 362 VRSEQEQRNLLDLLSTWTVPQI 385
Db 827 LAERQDRVDVFEILDWOLTKNV 850

RESULT 15
RRPO_PVMR STANDARD; PRT; 1968 AA.
ID PI7965; Q89548;
DT 01-NOV-1990 (Rel. 16, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA replication protein (147 kDa protein) (ORF 1) [Contains: RNA-
DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus M (strain Russian) (PVM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;
OC Carlavirus.
OX NCBI_TaxID=12168;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9116326; PubMed=1990070;
RA Zavrjev S.K., Kanyuka K.V., Levay K.E.;
RT "The genome organization of potato virus M RNA."
RL J. Gen. Virol. 72:9-14(1991).
CC -!- FUNCTION: RNA-replication. The central part of this protein
CC possibly functions as an ATP-binding helicase.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -!- SIMILARITY: Contains 1 OTU domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14449; BAA03339.1; -
CC PIR; P00093; PNO093. -
CC MEROPS; C23.001; -
CC InterPro; IPR003323; OTU.
CC InterPro; IPR008041; Peptidase_C23.
CC InterPro; IPR001788; RNA_dep_RNapol2.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC InterPro; IPR006006; Viral_helicase1.
CC Pfam; PF02338; OTU; 1.
CC Pfam; PF05379; Peptidase_C23; 1.
CC Pfam; PF00978; RNA_dep_RNapol2; 1.
CC Pfam; PF01443; Viral_helicase1; 1.
CC PROSITE; PS50802; OTU; 1.
CC ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
CC Transferase.
CC DOMAIN 883 991 OTU.
CC NP_BIND 1166 1173 ATP (POTENTIAL)
CC SEQUENCE 1968 AA; 223384 MW; 6F15A79E1AD96AAC CRC64;

Query Match 4.4%; Score 89; DB 1; Length 1968;
Best Local Similarity 20.7%; Pred. No. 36;
Matches 71; Conservative 38; Mismatches 108; Indels 126; Gaps 18;

QY 23 FFHYSHNTSSLKMQAPRMHVLVLSWRSGSSFVQGLFGQHPDV----- 68
```

Db 367 FARUVIHN-SMCAITIMEQL-----KEFMGNWLGKMPSVLARRFSSVRACVN 414
QY 69 -----FYLMEPAWHVMTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPG 116
Db 415 KFIRGLKEYSFTLRLEITW--WNIWENSYAWF-----FDTDAEVD 453
QY 117 -PRROSSLFQWENS---RALCSAPACDIIPODEIIPRAHCELLC--SQOPEVVEK---- 166
Db 454 VPEKLDLSLFMGEGAGLVAHITSRPYVGTVP---LADREWNALLCWDOSQKLLHAMRMMR 510
QY 167 -ACRSYSHVVLKEVRFFNLQSLYPLLKDPSSLNLHIVLVRDPRAVFRSRERTKGLMIDS 225
Db 511 GAWGAHMCVISRE---FLKXVEARLKSSCL-----IAKARR----- 545
QY 226 RIVMGOHEQK-----IKKEDQPYVMQVICOSQLEIYKTIQSLPKALQERYLLVRYE 277
Db 546 ----GOHKEKLEAWEVGLGKSSDALFRAMTYLCNARLEPMFSESGL-----RFFLTRGR 595
QY 278 DLRAPVAQTSRMVEFVGLFPLHLOTW---VHNIT--RGKGM 315
Db 596 NNLYGLTNYTEGKRAVTGVQNL-----MSNVVHEVSTKRHKGM 633

Search completed: May 6, 2004, 10:50:27
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:47:22 ; Search time 44 Seconds
(without alignments)
2767.955 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKMKLLFLVSQL.....EQRNLLDLLSTWTVPEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_prodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	100.0	386	Q9Y3R3	Q9Y3R3 homo sapien
2	2035	99.9	386	Q8NCG5	Q8NCG5 homo sapien
3	1963	96.3	370	Q8IV46	Q8IV46 homo sapien
4	1515.5	74.4	388	Q9X1I1	Q9X1I1 mus musculus
5	1513.5	74.3	388	Q9WUE5	Q9WUE5 mus musculus
6	1019	50.0	395	Q8GXZ3	Q8GXZ3 homo sapien
7	1017.5	49.9	395	Q9QUP4	Q9QUP4 mus musculus
8	1008	49.5	411	Q9GZS9	Q9GZS9 homo sapien
9	627	30.8	484	Q9EP78	Q9EP78 mus musculus
10	624	30.6	484	Q99NB0	Q99NB0 mus musculus
11	606.5	29.8	486	Q75667	Q75667 homo sapien
12	605.5	29.7	486	Q9NS94	Q9NS94 homo sapien
13	588.5	29.4	530	Q88276	Q88276 mus musculus
14	595.5	29.2	530	Q80WV3	Q80WV3 mus musculus
15	591.5	29.0	483	Q9UED5	Q9UED5 homo sapien
16	591.5	29.0	530	Q9Y4C5	Q9Y4C5 homo sapien

17	559	27.4	411	11	Q9EQC0	Q9eqc0 mus musculu
18	549	26.9	411	4	Q43916	Q43916 homo sapien
19	504.5	24.8	472	11	Q88199	Q88199 mus musculu
20	482	23.7	479	4	Q75099	Q75099 homo sapien
21	481.5	23.6	474	11	Q9QZL2	Q9qz12 rattus norv
22	423.5	20.8	441	13	Q83403	Q93403 torpedo cal
23	264	13.0	486	5	Q9VMC3	Q9vmc3 drosophila
24	215	10.5	120	11	Q9DOK5	Q9d0k5 mus musculu
25	210	10.3	363	5	Q9VMC4	Q9vmc4 drosophila
26	199	9.8	363	5	Q8MZD1	Q8mzd1 drosophila
27	175	8.6	119	6	Q95J48	Q95j48 oryctolagus
28	168	8.2	307	16	Q92VG4	Q92vg4 rhizobium m
29	124.5	6.1	274	16	Q8PRA0	Q8pra0 xanthomonas
30	118	5.8	368	16	Q8AAP8	Q8aap8 bacteroides
31	116.5	5.7	413	16	Q7UP52	Q7up52 rhodospirell
32	115.5	5.7	303	16	Q93JE5	Q93je6 streptomyce
33	111	5.6	360	5	Q86R90	Q86r90 halocynthia
34	111	5.4	655	6	Q9BX00	Q9bx00 macaca fasc
35	108.5	5.3	1222	4	Q8IZU8	Q8izu8 homo sapien
36	107	5.3	1586	4	Q9P2D3	Q9p2d3 homo sapien
37	100.5	4.9	747	4	Q96QU2	Q96qu2 homo sapien
38	100.5	4.9	762	17	Q8OCG7	Q8ocg7 methanosarc
39	100.5	4.9	775	4	Q9C011	Q9c011 homo sapien
40	100	4.9	593	10	Q8VZE0	Q8vze0 arabidopsis
41	100	4.9	593	10	Q8ARR2	Q8arr2 arabidopsis
42	97.5	4.8	285	2	Q8KWX0	Q8kwx0 xanthomonas
43	96.5	4.7	743	4	Q9H0K2	Q9h0k2 homo sapien
44	96	4.7	2454	3	Q9UV56	Q9uv56 emericeila
45	96	4.7	2454	3	Q9UVP2	Q9uvp2 emericeila

ALIGNMENTS

RESULT 1

Q9Y5R3 PRELIMINARY; PRT; 386 AA.

AC Q9Y5R3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (U-selectin ligand
DE sulfotransferase GST-3).
GN GST3.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A.

RP TISSUE=Tonsil;

RX MEDLINE=92264336; PubMed=10330415;

RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.R.,

RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;

RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).

[2] SEQUENCE FROM N.A.

RP TISSUE=Tonsil;

RC Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,

RA Rosen S.D.;

RT "Chromosomal Localization and Genomic Organization for the

RT Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-

RT Sulfotransferase Gene Family.";

RN [3] Glycobiology 0:0-0(2001).

RP SEQUENCE FROM N.A.

RM MEDLINE=21332592; PubMed=11439191;

RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellies L.G.,

RA Rabuka D., Hindsaun O., Marth J.D., Lowe J.B., Fukuda M.;

RT "Novel sulfated lymphocyte homing receptors and their control by a

RT core1 extension beta1,3-N-acetylglucosaminyltransferase.";

Cell 105:957-969 (2001).
 RL EMBL; AF131235; AD33015.1; -;
 DR EMBL; AF280088; AK48246.1; -;
 DR EMBL; AF149783; AK48417.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0006928; P:cell motility; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0006955; P:immune response; TAS.
 DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
 DR InterPro; IPR001092; HLH basic.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 KW Lectin; Selectin; Transferase.
 SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 100.0%; Score 2038; DB 4; Length 386;
 Best Local Similarity 100.0%; Pred. No. 5.5e-183; Indels 0; Gaps 0;
 Matches 386; Conservative 0; Mismatches 0;

QY 1 MLLPKKKLLFLVLSQMAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQ 60
 DB 1 MLLPKKKLLFLVLSQMAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQ 60
 QY 61 LFGQHPDVFYLMPEPAHVMWTFKOSTAWMLHMAVRLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFYLMPEPAHVMWTFKOSTAWMLHMAVRLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSIFQWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 DB 121 SSIFQWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 QY 181 FFNLQSLYPLKDPKPSLNLIHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 DB 181 FFNLQSLYPLKDPKPSLNLIHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 QY 241 QPYVQVVCOSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 DB 241 QPYVQVVCOSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 QY 301 HLQTVHNITRGKMGDGHAFHTNARDALNVSAQWNSLPYKVSRLQKACGDANNLLGYR 360
 DB 301 HLQTVHNITRGKMGDGHAFHTNARDALNVSAQWNSLPYKVSRLQKACGDANNLLGYR 360
 QY 361 HVSEQEQRNLLDLSTWTVPQIH 386
 DB 361 HVSEQEQRNLLDLSTWTVPQIH 386

RESULT 2
 Q8NCG5 PRELIMINARY; PRT; 386 AA.
 AC Q8NCG5
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein FLJ30265.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC035282; AAH35282.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR GO; GO:0016740; P:transferase activity; IEA.
 DR InterPro; IPR001092; HLH basic.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 KW Transferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 96.3%; Score 1963; DB 4; Length 370;
 Best Local Similarity 100.0%; Pred. No. 5.9e-176;
 Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQLFGQHPDVFYLMPEPAW 76
 DB 17 MAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQLFGQHPDVFYLMPEPAW 76

DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 99.9%; Score 2035; DB 4; Length 386;
 Best Local Similarity 99.7%; Pred. No. 1.1e-182;
 Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPKKKLLFLVLSQMAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQ 60
 DB 1 MLLPKKKLLFLVLSQMAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQ 60
 QY 61 LFGQHPDVFYLMPEPAHVMWTFKOSTAWMLHMAVRLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFYLMPEPAHVMWTFKOSTAWMLHMAVRLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSIFQWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 DB 121 SSIFQWNSRALCSAPACDIIPODEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVYLKEVR 180
 QY 181 FFNLQSLYPLKDPKPSLNLIHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 DB 181 FFNLQSLYPLKDPKPSLNLIHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
 QY 241 QPYVQVVCOSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 DB 241 QPYVQVVCOSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGLFELP 300
 QY 301 HLQTVHNITRGKMGDGHAFHTNARDALNVSAQWNSLPYKVSRLQKACGDANNLLGYR 360
 DB 301 HLQTVHNITRGKMGDGHAFHTNARDALNVSAQWNSLPYKVSRLQKACGDANNLLGYR 360
 QY 361 HVSEQEQRNLLDLSTWTVPQIH 386
 DB 361 HVSEQEQRNLLDLSTWTVPQIH 386

RESULT 3
 Q8IV46 PRELIMINARY; PRT; 370 AA.
 AC Q8IV46
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC035282; AAH35282.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR GO; GO:0016740; P:transferase activity; IEA.
 DR InterPro; IPR001092; HLH basic.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 KW Transferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 96.3%; Score 1963; DB 4; Length 370;
 Best Local Similarity 100.0%; Pred. No. 5.9e-176;
 Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQLFGQHPDVFYLMPEPAW 76
 DB 17 MAILAFFHMYSHNISLSMKQAQPERMHVLSVSRSGSSFVGQLFGQHPDVFYLMPEPAW 76

```
Db 1 MAILALPFMYSHNISLSMKAQPERMHVILVLSWRSGSFVGLQFGHPDVFYLMPEAW 60
QY 77 HVWTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRROSSLFQWENSRLCSAP 136
Db 61 HVWTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPRROSSLFQWENSRLCSAP 120
QY 137 ACDIIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVFFNLQSLYPLKDPSSL 196
Db 121 ACDIIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVFFNLQSLYPLKDPSSL 180
QY 197 NLHIVHLVDRPRAVRSRRTKGLMDSRVIMGHEOKLAKEDQPYVWQVVCOSOLEI 256
Db 161 NLHIVHLVDRPRAVRSRRTKGLMDSRVIMGHEOKLAKEDQPYVWQVVCOSOLEI 240
QY 257 YKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFELPHLQTTWVHNITRGKMG 316
Db 241 YKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFELPHLQTTWVHNITRGKMG 300
QY 317 DHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGYRHRVSEQQRNLLDLL 376
Db 301 DHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGYRHRVSEQQRNLLDLL 360
QY 377 STWTVPEQIH 386
Db 361 STWTVPEQIH 370
RESULT 4
Q9R111 ID Q9R111 PRELIMINARY; PRT; 388 AA.
AC Q9R111
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE L-selectin ligand sulfotransferase.
GN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93361934; PubMed=10435581;
RA Hiraoka N., Petryniak B., Nakayama J., Teubol S., Suzuki M., Yeh J.C.,
RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34."
RL Immunity 11:79-89(1999).
DR EMBL; AF109155; AAD45579.1; -.
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:000529; F:sugar binding; IEA.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferrase activity; IEA.
DR GO; GO:0007157; P:heterophilic cell adhesion; IEA.
DR GO; GO:000685; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR Lectin; Selectin; Transferrase.
SQ SEQUENCE 388 AA; 44636 MW; 6D5371AFB6884AEE CRC64;
Query Match 74.4%; Score 1515.5; DB 11; Length 388;
Best Local Similarity 72.8%; Pred. No. 7.1e-134;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;
QY 1 MLLPKXMKLLLFVQSMAILALFFHMYSHNISLSMKAQPER-MHYVLVLSWRSGSFVYG 59
Db 1 MMLKKGRLLMFLGQVIVVALFIHMSVHR-HLSQREBSRRPVHVLVLSWRSGSFVG 58
QY 60 QLFQGHDPDVFYLMPEAWHVWTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPR 119
Db 59 QLFQGHDPDVFYLMPEAWHVWTFKOSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPR 118
QY 120 QSSILFQWENSRLCSAPACDIIIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEV 179
```

```
Db 119 QSSILFQWENSRLCSAPACDIIIPQDEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEV 178
QY 180 RFENLQSLYPLKDPSSLNLHIVHLVDRPRAVRSRRTKGLMDSRVIMGHEOKLAKKE 239
Db 179 RFENLQSLYPLKDPSSLNLHIVHLVDRPRAVRSRRTKGLMDSRVIMGHEOKLAKKE 238
QY 240 DQPYVWQVVCOSOLEIYKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFEL 299
Db 239 DQPYVWQVVCOSOLEIYKTIQSLPKALQERYLLVRYEDLAPVAQTSRMVYFVGLFEL 298
QY 300 PHLQTTWVHNITRGKMGDHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGY 359
Db 299 PHLQTTWVHNITRGKMGDHAFHTNARDALNVQAWRWSLPYEKVSRLQKACGDAMNLLGY 358
QY 360 RHYRSEQQRNLLDLLSTWTVPEQI 385
Db 359 LQVRSQEQGNLSLDLLSSSHILGOV 384
RESULT 5
Q9WUE5 ID Q9WUE5 PRELIMINARY; PRT; 388 AA.
AC Q9WUE5
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin
DE 6/keratan) sulfotransferase 4).
GN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Tonsil;
RX MEDLINE=9264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,
RA Huang C.-C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin."
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Hono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boifelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald C., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AF131236; AAD33016.1; -.
DR EMBL; AK009113; BAB26078.1; -.
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferrase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR InterPro; IPR000863; Sulfotransferase.
```


RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kuchiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wolming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [4]
RN SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RP Strausberg R.,
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280089; AAG48247.1; -;
DR EMBL; AK040710; BAB13769.1; -;
DR EMBL; AK011202; BAB27465.1; -;
DR EMBL; BC019204; AAH19204.1; -;
DR PIR; JC7350; JC7350.
DR MGD; MGI:1891767; Chat7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 484 AA; 54766 MW; 98195537D7AB7193 CRC64;

Query Match 30.8%; Score 627; DB 11; Length 484;
Best Local Similarity 34.4%; Pred. No. 3e-50;
Matches 137; Conservative 65; Mismatches 150; Indels 46; Gaps 7;

QY 30 NISSLSMKAQPERMHVLLSSWRSGSFVGLFGHPDVFYLMPEAHVMMTFKQSTAWM 89
DB 87 NLSAVEAVTQEKHIYVHATWRTGSSFLGELFNQHPDVFYLMPEAHVMMTFKQSTAWM 146
QY 90 LHMVRLIRAVFLCDMSVFDAYMEPPGPRQ-----SSLPQWENSRLCSAPACD 139
DB 147 LQALRDLRLSLFRCDVSLRYAQDPGERAPDSANLTAMLFRWRTNKVICSPPLCP 206
QY 140 IIPQDE----IIPRAHCRLLCSQPFVEVVEKACRSYSHWLKEVRFNQLSLYPLLKDPS 195
DB 207 AAPRADVGLVEDKACESTCPVSLRALEAECKYPVVVVKDVLGLVPLLRDPG 266
QY 140 IIPQDE----IIPRAHCRLLCSQPFVEVVEKACRSYSHWLKEVRFNQLSLYPLLKDPS 195
DB 207 AAPRADVGLVEDKACESTCPVSLRALEAECKYPVVVVKDVLGLVPLLRDPG 266
QY 196 LNLHIVLRDPRVFRSRTKGLMIDS-----RIVM-----GQHE 233
DB 267 LNLKVVQLFRDPRVHNSRLKSQGLLRRESIQVLRQGRGHFHVLLAHGVDPGQQA 326
QY 234 QKLKEDQPY-----VMQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVQTSR 289
DB 327 RALPSAPRADFFLTSALEVICEAWLRDLFTRGAPAWLRRLRYEDLVWQQAQLRR 386
QY 290 MYEFVGLLEFLPLQTVWNITRGKMG-DHAFHTNARDALNVSAQRWSLPYEKYSRLQK 348
DB 387 LLRFSGRLTLAALDAFAFNTRGSAYGADRPFLHSARDAREAVHAWRRLSQEQVRQVET 446
QY 349 ACGDAMNLLGYHRVSRSEQQRNLL-----LDLLSTWTV 381
DB 447 ACAPAMRLLAYPRSGDERDKTVREGTFLETKANNAV 484

RESULT 10
Q99NB0 PRELIMINARY; PRT; 484 AA.
AC Q99NB0;

DT 01-JUN-2001 (T-REMBLrel. 17, Created)
DT 01-JUN-2001 (T-REMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE Chondroitin 6-sulfotransferase-2.
GN CHST7 OR MC6ST-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Kitagawa H., Uyama T., Sugahara K.,
RT "Cloning and Expression of Mouse Chondroitin 6-sulfotransferase-2."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046529; BAB40372.1; -;
DR PIR; JC7350; JC7350.
DR MGD; MGI:1891767; Chat7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 484 AA; 54751 MW; 402C5E1ED185PDF8 CRC64;

Query Match 30.6%; Score 624; DB 11; Length 484;
Best Local Similarity 34.2%; Pred. No. 5.8e-50;
Matches 136; Conservative 66; Mismatches 150; Indels 46; Gaps 7;

QY 30 NISSLSMKAQPERMHVLLSSWRSGSFVGLFGHPDVFYLMPEAHVMMTFKQSTAWM 89
DB 87 NLSAVEAVTQEKHIYVHATWRTGSSFLGELFNQHPDVFYLMPEAHVMMTFKQSTAWM 146
QY 90 LHMVRLIRAVFLCDMSVFDAYMEPPGPRQ-----SSLPQWENSRLCSAPACD 139
DB 147 LQALRDLRLSLFRCDVSLRYAQDPGERAPDSANLTAMLFRWRTNKVICSPPLCP 206
QY 140 IIPQDE----IIPRAHCRLLCSQPFVEVVEKACRSYSHWLKEVRFNQLSLYPLLKDPS 195
DB 207 AAPRADVGLVEDKACESTCPVSLRALEAECKYPVVVVKDVLGLVPLLRDPG 266
QY 196 LNLHIVLRDPRVFRSRTKGLMIDS-----RIVM-----GQHE 233
DB 267 LNLKVVQLFRDPRVHNSRLKSQGLLRRESIQVLRQGRGHFHVLLAHGVDPGQQA 326
QY 234 QKLKEDQPY-----VMQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVQTSR 289
DB 327 RALPSAPRADFFLTSALEVICEAWLRDLFTRGAPAWLRRLRYEDLVWQQAQLRR 386
QY 290 MYEFVGLLEFLPLQTVWNITRGKMG-DHAFHTNARDALNVSAQRWSLPYEKYSRLQK 348
DB 387 LLRFSGRLTLAALDAFAFNTRGSAYGADRPFLHSARDAREAVHAWRRLSQEQVRQVET 446
QY 349 ACGDAMNLLGYHRVSRSEQQRNLL-----LDLLSTWTV 381
DB 447 ACAPAMRLLAYPRSGDERDKTVREGTFLETKANNAV 484

RESULT 11
Q75667 PRELIMINARY; PRT; 486 AA.
AC Q75667;
DT 01-NOV-1998 (T-REMBLrel. 08, Created)
DT 01-NOV-1998 (T-REMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (Carbohydrate
DE (N-acetylglucosamine 6-O) sulfotransferase 7).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Brain;
RA Uchimura K., Muramatsu T.;
RT "Identification and molecular characterization of a cDNA encoding a
RT novel N-acetylglucosamine-6-O-sulfotransferase.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL EMBL; AB040711; BAB13770.1; -;
RL EMBL; BC045537; AAH45537.1; -;
DR FR; JC7351; JC7351.
DR Genew; HGNC:13817; CHST7.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 486 AA; 54266 MW; 3F1FD1430B3C8E95 CRC64;
Query Match 29.8%; Score 606.5; DB 4; Length 486;
Best Local Similarity 37.0%; Pred. No. 2.6e-48;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;
QY 41 ERMHVLVLSWRSGSFVQLFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRA 100
DB 100 EKQHIYHATWRTGSSFLGELFNQHPDVFYLYEPMMHLWQALYPGDAESLQGLRDMLS 159
QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SSLFQWENSRLCSAPACDIIPQDE---I 146
DB 160 LFRCDFSVLRLYAPPGDPAARAPDTANLTAAALFRWRTNKVICSPPLCPGAPRAAEVGL 219
QY 147 IPRAHCRLLCSQQPFVEWEKACBSYSHVLEKVRFPNLSQLYPLPKDPSLNHLHVLVRD 206
DB 220 VEDTACERSCPPVAIRALEAECKYPVWVIKDVRLDLDGLVFLRLDPGLNLKVQLFRD 279
QY 207 PRAVFRSRETKGDLMTDS-----RIVM-----GOHEQKLKKEQOPY 244
DB 280 PRAVHNSRLKSROGLLRRESIQVLRTRQGRDFHRLVLAHGVGARGPGQSRALPAAPRADF 339
QY 245 -----VMQVICOQLEIKYKTIQSLPKALQERYLLRVYEDLARPAVACTSRMYEFVGLFLP 300
DB 340 FLTGALVEICEAWRLDLPFARGAPAWLRRLRYEDLVROPQALRLRLRFSGLRALA 399
QY 301 HLQTVHNITRGKMG-DAHFHTNARDALNVSAWMSLPYKVSRLQKACGDAMNLLGY 359
DB 400 ALDAFALNMTGGAAYGADRPFLHSARDAREAVHWRERLSRQVROVEVACAPAMLLAY 459
QY 360 RHVRSEQE 367
DB 460 --PRSGEE 465

RESULT 12

Q9NS84 PRELIMINARY; PRT; 486 AA.
AC Q9NS84;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Chondroitin 6-sulfotransferase-2.
GN CSST-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20347214; PubMed=10781596;
RA Kitagawa H., Fujita M., Ito N., Sugahara K.;
RT "Molecular cloning and expression of a novel chondroitin 6-O-

RT sulfotransferase.";
RL J. Biol. Chem. 275:21075-21080(2000).
DR EMBL; AB037187; BAB03217.1; -;
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0008459; F:chondroitin 6-sulfotransferase activity; TAS.
DR GO; GO:0005976; P:polysaccharide metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 486 AA; 54294 MW; B524D1430B3C9075 CRC64;
Query Match 29.7%; Score 605.5; DB 4; Length 486;
Best Local Similarity 37.0%; Pred. No. 3.2e-48;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;
QY 41 ERMHVLVLSWRSGSFVQLFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRA 100
DB 100 EKQHIYHATWRTGSSFLGELFNQHPDVFYLYEPMMHLWQALYPGDAESLQGLRDMLS 159
QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SSLFQWENSRLCSAPACDIIPQDE---I 146
DB 160 LFRCDFSVLRLYAPPGDPAARAPDTANLTAAALFRWRTNKVICSPPLCPGAPRAAEVGL 219
QY 147 IPRAHCRLLCSQQPFVEWEKACBSYSHVLEKVRFPNLSQLYPLPKDPSLNHLHVLVRD 206
DB 220 VEDTACERSCPPVAIRALEAECKYPVWVIKDVRLDLDGLVFLRLDPGLNLKVQLFRD 279
QY 207 PRAVFRSRETKGDLMTDS-----RIVM-----GOHEQKLKKEQOPY 244
DB 280 PRAVHNSRLKSROGLLRRESIQVLRTRQGRDFHRLVLAHGVGARGPGQSRALPAAPRADF 339
QY 245 -----VMQVICOQLEIKYKTIQSLPKALQERYLLRVYEDLARPAVACTSRMYEFVGLFLP 300
DB 340 FLTGALVEICEAWRLDLPFARGAPAWLRRLRYEDLVROPQALRLRLRFSGLRALA 399
QY 301 HLQTVHNITRGKMG-DAHFHTNARDALNVSAWMSLPYKVSRLQKACGDAMNLLGY 359
DB 400 ALDAFALNMTGGAAYGADRPFLHSARDAREAVHWRERLSRQVROVEVACAPAMLLAY 459
QY 360 RHVRSEQE 367
DB 460 --PRSGEE 465
RESULT 13
O88276 PRELIMINARY; PRT; 530 AA.
AC O88276;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN CHST2
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
EX MEDLINE=98380482; PubMed=9712885;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583(1998).
DR EMBL; AB011452; BAA32138.1; -;
DR EMBL; AB011452; BAA32139.1; -;
DR EMBL; AB011451; BAA32137.1; -;
DR MGB; MGI:1891160; Chst2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:45:57 ; Search time 58 Seconds
(without alignments)
1880.402 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLPLKKMLLFLVQMAIL.....PORNLLDLLSTWTVPQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2038	100.0	386	2 AAY39918	Aay39918 Human gly
2	2035	99.9	386	4 AAM93309	Aam93309 Human pol
3	2008	98.5	380	5 AAU11274	Aau11274 Human L-s
4	1936	95.0	386	3 AAY79219	Aay79219 Human tra
5	1513.5	74.3	388	2 AAY39919	Aay39919 Mouse gly
6	1205.5	59.2	418	3 AAB41947	Aab41947 Human ORF
7	1028	50.4	418	5 ABB81557	Abb81557 Mouse int
8	1019	50.0	395	4 AAY72640	Aay72640 Human gly
9	1019	50.0	395	5 ABB81554	Abb81554 Human cor
10	1019	50.0	395	5 AAE15438	Aae15438 Human dru
11	1017.5	49.9	395	5 AAY72638	Aay72638 Mouse gly
12	1017.5	49.9	395	5 AAU11275	Aau11275 Murine in
13	1013.5	49.7	395	5 ABB81555	Abb81555 Consensus
14	1008	49.5	390	4 AAY72639	Aay72639 Human gly
15	1008	49.5	390	5 ABB81556	Abb81556 Human int
16	821	40.3	171	5 ABB81560	Abb81560 Human hig
17	605.5	29.7	486	6 ABP56121	Abp56121 Human cho
18	598.5	29.4	483	3 AAY31656	Aay31656 Mouse N-a
19	591.5	29.0	530	4 AAB95367	Aab95367 Human pro
20	587.5	28.8	484	2 AAY31657	Aay31657 Human N-a
21	587.5	28.8	531	5 AAU69414	Aau69414 Lung smal
22	549	26.9	411	2 AAU61100	Aau61100 Keratan s
23	549	26.9	411	5 AAE25356	Aae25356 Human cho
24	549	26.9	411	6 ABU03503	Abu03503 Angiogene
25	519.5	25.5	169	5 ABB81558	Abb81558 Human cor

26	511.5	25.1	169	5	ABB81559	Abb81559 Human int
27	500.5	24.6	458	2	AAW06480	Aaw06480 Chick cho
28	492	24.1	499	6	ABR41139	Abra1139 Human DIT
29	482	23.7	479	2	AAW52883	Aaw52883 Glycosami
30	336	16.5	169	5	ABB81561	Abb81561 Human N-a
31	335.5	16.5	179	5	ABB81562	Abb81562 Human ker
32	293	14.4	174	5	ABB81563	Abb81563 Human cho
33	210	10.3	363	4	ABG64512	Abg64512 Drosophil
34	186.5	9.2	183	4	ABG65852	Abg65852 Mycobacte
35	116	5.7	388	6	AAE32782	Aae32782 Mouse gly
36	109	5.3	1207	4	AAU72643	Aau72643 Human nov
37	109	5.3	1207	6	AAE33542	Aae33542 Human gly
38	108.5	5.3	596	4	AAU72641	Aau72641 Human nov
39	108.5	5.3	1212	6	AAE33541	Aae33541 Human nov
40	108.5	5.3	1222	4	AAU72642	Aau72642 Human gly
41	108.5	5.3	1222	6	ABU11849	Abu11849 Human sec
42	100.5	4.9	775	5	ABB97942	Abb97942 Human pro
43	100	4.9	593	7	ADD30841	Add30841 Plant yie
44	100	4.9	593	7	ADE31489	Ade31489 Plant yie
45	96.5	4.7	743	4	ABU52986	Abu52986 Human nuc

ALIGNMENTS

RESULT 1
AAY39918
ID AAY39918 standard; protein; 386 AA.

XX AAY39918;
AC AAY39918;
XX 08-DEC-1999 (first entry)

DE Human glycosyl sulfotransferase-3 protein sequence.

XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;

KW selectin binding interaction; inflammation; lymphocyte homing; human;

KW secondary lymph organ.

XX Homo sapiens.

XX WO9949018-A1.

XX 30-SEP-1999.

XX 26-FEB-1999; 99WO-US004316.

XX 20-MAR-1998; 98US-00045284.

PR 12-NOV-1998; 98US-00190911.

XX (REGC) UNIV CALIFORNIA.

XX (SYNT) SYNTX USA INC.

XX Bistrup A, Rosen SD, Targemann K, Hemmerich S;

XX WPI; 1999-580442/49.

XX N-PSDB; AAZ20792.

XX Human and murine glycosyl sulfotransferase 3 and related polynucleotides.

XX Claim 2; Fig 1; 59pp; English.

XX This sequence is the human glycosyl sulfotransferase-3 (GST-3) of the

XX invention. The nucleic acid sequences, probes and primers derived from

XX these, proteins and antibodies are useful in detecting homologues. The

XX sequences, antibodies and methods are useful in the diagnosis and

XX treatment of diseases associated with selectin binding interactions,

XX including conditions associated with or resulting from the homing of

XX leukocytes to sites of inflammation and the normal homing of lymphocytes

XX to secondary lymph organs

XX Sequence 386 AA;

PA (BURN-) BURNHAM INST.
 XX
 PI Fukuda M, Yeh J, Hiraoka N;
 XX
 DR WPI; 2002-075226/10.
 DR N-PSDB; AAS16947.
 XX
 XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
 PT L-selectin sulfotransferase-2 that directs expression of L-selectin
 PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
 PT GLCNAc 6-sulfotransferase.
 XX
 PS Claim 21; Fig 4; 98pp; English.
 XX
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated betal, 3-N-
 CC acetylglucosaminyltransferase (betal, 3GnT) or an active fragment, where
 CC betal, 3GnT directs expression of a MECA-79 antigen. The invention also
 CC provides a method of treating or preventing an L-selectin-mediated
 CC condition by reducing the expression or activity of a betal, 3GnT that
 CC directs expression of a MECA-79 antigen. This can be done by
 CC administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC and/or a betal, 3GnT antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of betal, 3GnT. The method is useful for treating the expression or
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents human LST-2
 XX
 XX Sequence 380 AA;
 SQ
 Query Match 98.5%; Score 2008; DB 5; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.6e-202;
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 MKLLFLVSMATLALFFHMYSHNTSSLSKKAQPERMHVLTSSWSGSSSVGQLFGQHP 66
 Db 1 MKLLFLVSMATLALFFHMYSHNTSSLSKKAQPERMHVLTSSWSGSSSVGQLFGQHP 60
 QY 67 DVFYLMPEAHVWMTFKQSTAWMLHNAVRLDLIRAVFLCDMSVFDAYMEFGPRRSSLFQW 125
 Db 61 DVFYLMPEAHVWMTFKQSTAWMLHNAVRLDLIRAVFLCDMSVFDAYMEFGPRRSSLFQW 120
 QY 127 ENSRALCSAPACDIIIPQDEIIIPRAHCRILCSQOPFVVEKACRSYSHVYLKEVRFNLOS 186
 Db 121 ENSRALCSAPACDIIIPQDEIIIPRAHCRILCSQOPFVVEKACRSYSHVYLKEVRFNLOS 180
 QY 187 LYPCLKDPSLNLHIVHLVDRPRAVFRSRETRTKGDLMDISRIVWGQHEQKLKKEQPPYVW 246
 Db 181 LYPCLKDPSLNLHIVHLVDRPRAVFRSRETRTKGDLMDISRIVWGQHEQKLKKEQPPYVW 240
 QY 247 QVICQSOLEYKTIQSLPQALQRYLLVRYEDLARAPVQTSRMVFEVGLBFLHLQVW 306
 Db 241 QVICQSOLEYKTIQSLPQALQRYLLVRYEDLARAPVQTSRMVFEVGLBFLHLQVW 300
 QY 307 HNTTRGKMGDGHAFHTNARDALNVSQAWWSLPYKVSRLQKACGDAMNLLGYRHRVSEQ 366
 Db 301 HNTTRGKMGDGHAFHTNARDALNVSQAWWSLPYKVSRLQKACGDAMNLLGYRHRVSEQ 360
 QY 367 EQRNLLDLLLSTWTVPEQIH 386
 Db 361 EQRNLLDLLLSTWTVPEQIH 380

RESULT 4
 AAY79219
 ID AAY79219 standard; protein; 386 AA.

XX
 AC
 XX
 DT 19-JUN-2000 (first entry)
 XX
 DE Human transferase TRNSFS-11.
 XX
 KW Transferase: TRNSFS-11; human; antitumor; cell proliferation;
 KW gastrointestinal disorder; developmental disorder; genetic disorder;
 KW neurological disorder; reproductive disorder; smooth muscle disorder;
 KW immunological disorder; inflammation; diagnosis; therapy;
 XX
 KW N-acetylglucosamine 6-O-sulfotransferase.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Domain 7..23 "transmembrane domain"
 FT Modified-site 30 /note= "potential N-glycosylation"
 FT Modified-site 35 /note= "potential O-phosphorylation"
 FT Modified-site 50 /note= "potential O-phosphorylation"
 FT Modified-site 81 /note= "potential O-phosphorylation"
 FT Modified-site 107 /note= "potential O-phosphorylation"
 FT Modified-site 121 /note= "potential O-phosphorylation"
 FT Modified-site 217 /note= "potential O-phosphorylation"
 FT Modified-site 243 /note= "potential O-phosphorylation"
 FT Modified-site 252 /note= "potential O-phosphorylation"
 FT Modified-site 287 /note= "potential O-phosphorylation"
 FT Modified-site 308 /note= "potential N-glycosylation"
 FT Modified-site 329 /note= "potential N-glycosylation"
 FT Modified-site 364 /note= "potential O-phosphorylation"
 FT Modified-site 380 /note= "potential O-phosphorylation"
 XX
 XX WO200014251-A2.

XX
 PD 16-MAR-2000.
 XX
 PF 09-SEP-1999; 99WO-US020989.
 XX
 PR 10-SEP-1998; 98US-00150657.
 PR 04-NOV-1998; 98US-00186779.
 PR 11-MAY-1999; 99US-0133642P.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
 PI Hillman JL, Azimzai Y;
 XX
 DR WPI; 2000-256996/22.
 DR N-PSDB; AA294211.
 XX
 XX Human transferase proteins useful for preventing, diagnosing and treating
 PT cancers and developmental, gastrointestinal, genetic, immunological,
 PT neurological, reproductive and smooth muscle disorders.
 XX
 PS Claim 1; Page 90-91; 113pp; English.
 XX
 CC The present sequence is that of human transferase TRNSFS-11, 1 of 15
 CC claimed novel human transferase proteins of the invention (see AAY79209-

23). The sequence was deduced from a cDNA clone (see A294211) isolated from a gallbladder library. It shows homology to mouse N-acetylglucosamine 6-C-sulfotransferase. RNS5-11 is expressed in dermatologic and gastrointestinal tissues, especially those associated with inflammation and cell proliferation. The new human transferrase and polynucleotides can be used in the diagnosis, prevention and treatment of cancer, developmental disorders, gastrointestinal disorders, genetic disorders, immunological disorders, neurological disorders, reproductive disorders, and smooth muscle disorders. The polypeptides can also be used to raise antibodies, and to screen for agonists and antagonists of transferase activity

SQ Sequence 386 AA; Query Match 95.0%; Score 1936; DB 3; Length 386;
 Best Local Similarity 95.6%; Pred. No. 6.4e-195;
 Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy	1	MLLPKOKULLLFLVSCMAILALFFHMYSHN1SSLSMKAQPERMHVLVLSWRSGSSFVQ	60
Db	1	MLLPKOKULLLFLVSCMAILALFFHMYSHN1SSLSMKAQPERMHVLVLSWRSGSSFVQ	60
Qy	61	LFQGHDPVYLMPEPAWHVMTPKQSTAMWLHMAVRDLIRAVFLCDMSVFDAYNEPQPRQ	120
Db	61	LFQGHDPVYLMPEPAWHVMTPKQSTAMWLHMAVRDLIRAVFLCDMSVFDAYNEPQPRQ	120
Qy	121	SSLFQWENSRALCSAPACDIIPQDEIIIPRAHCRLLCSQQPFVEVKEACRSYSHVVLKEVR	180
Db	121	SSLFQWENSRALCSAPACDIIPQDESSPGLTAGSCAVNSSLKLEKACRSYSHVVLKEVR	180
Qy	181	FFNLQSLYPLLDPSLNLIHIVHVRDPRAVFRSBERTKGDLMDSRIVMGQHQKLLKED	240
Db	181	FFNLQSLYPLLDPSLNLIHIVHVRDPRAVFRSBERTKGDLMDSRIVMGQHQKLLKED	240
Qy	241	QPIYVQVVCOSOLEIYKTIQSLPRALQERYLLVRYEDLARAPVAQTSRMVEFVGLFEFLP	300
Db	241	QPIYVQVVCOSOLEIYKTIQSLPRALQERYLLVRYEDLARAPVAQTSRMVEFVGLFEFLP	300
Qy	301	HLQTTWHNITRKGMDGHAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNLLGYR	360
Db	301	HLQTTWHNITRKGMDGHAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNLLGYR	360
Qy	361	HVSEQEQORNLLDLLSTWTVPQIH	386
Db	361	HVSEQEQORNLLDLLSTWTVPQIH	386

RESULT 5
AAY39919
ID AAY39919 standard: protein: 388 AA.

AA
AC
AAAY39919:

DT 08-DEC-1999 (first entry)

XX
DE Mouse glycosyl sulfotransferase-3 protein sequence.

XX Glycyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.

XX
OS
Mus sp.

XX
PN WO9949018-A1.

XX 30-SEP-

26-FEB-1999: 99WO-US004316.

XX
PP 20-MAR-1998. 98IIS-00045284

PR 12-NOV-1998; 98US-00190911.
yy

PA (REGC) UNIV CALIFORNIA.

(SYNT) SYNTAX USA INC.

Bistrup A, Rosen SD, Tangemann K, Hemmerich S;

WPI: 1999-580442/49.

N-PSUB; AA720133:

Human and murine glycosyl sulfolactonase 3 and related F₀/F₁ membrane proteins

Claim 2: Fig 4: 59pp: English.

This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.

Sequence 388 AA:

74.3%: Score 1513.5; DB 2; Length 388;

st Local Similarity	72.8%;	Pred: NO: 2:28-150;	
atches	281:	Conservative	49: Mismatches
			53: Indels
			3: Gaps

1 MT.I.PKKMKLLLELVVSOMAILALFFHMYSHNISSLSMKAQPER-MHVLVLSSWRSGSSFVG 59

1 MMTLKKCPRIIMETLGSQVTVVAIFTHMSVHR--HLSOREESRRPVHVLVLSWRSGSSFVG 58

60 01 EGOHBDVEYI MEPAWHVWTEKOSTAWMIHMAVRDLIRAVELCDMSVFDA YMEPGRR 119

07 PQOJNDNVEYI MEDA WLTZMTETSSTAWKIHMAVRDILRSVFTCDMSVEDAYMNPGRK 118

ZOO OGGT BOGTHNGDNY CCGA BZAGCSTI DQDETT DBXKCBPILJCSOOPBFEVWEKA CRSYSHVL KEV 179

[illegible]

100 239

238

[illegible]

298

----- 359

[illegible][illegible]

Detailed description of the gel electrophoresis image: The image shows a vertical gel with 10 lanes. Lane 1 contains a DNA ladder with multiple bands of varying sizes. Lanes 2 through 10 contain DNA samples. Lane 2 has a single band marked with an asterisk (*). Lane 3 has a single band marked with an asterisk (*). Lane 4 has a single band marked with an asterisk (*). Lane 5 has a single band marked with an asterisk (*). Lane 6 has a single band marked with an asterisk (*). Lane 7 has a single band marked with an asterisk (*). Lane 8 has a single band marked with an asterisk (*). Lane 9 has a single band marked with an asterisk (*). Lane 10 has a single band marked with an asterisk (*).

RESULT 6

AAB41947

ID AAB41947 standard; protein; 418 AA.

AC AAB41947;

08-FEB-2001 (first entry)

Human OREF ORF1711 polypeptide sequence SEQ ID NO:3422.

XX Hunan; open reading frame; ORF; detection; cytostatic; hepatotropic;
KW anti-hepatitis; anti-hepatitic; anti-hepatinspan; necrotic; neuroprotective;
KW anti-hepatitis; anti-hepatitic; anti-hepatinspan; necrotic; neuroprotective;
KW immunosuppressant; osteopathic; articular; immunosuppressant; cardiac;
KW immunosuppressant; osteopathic; articular; immunosuppressant; cardiac;
KW hypotensive; dermatological; immunosuppressive; anti-inflammatory;
KW antiviral; antibacterial; antifungal; anti-heumatic; anti-thyroid;
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW

PA
VY

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
FN WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC76156.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
FT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 2599-2600; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 418 AA;
Query Match 59.2%; Score 1205.5; DB 3; Length 418;
Best Local Similarity 61.2%; Pred. No. 7.1e-118;
Matches 241; Conservative 47; Mismatches 79; Indels 27; Gaps 6;
QY 1 MLLPKXKLLFLVNSQMAILALFFHMYSHNIISSLSMKAKQPERMHVLSVSRSGSFVQG 60
41 LLLAQTTCLLIISRP-----GFSPPAGGEDAVHVLSSWRSGSFLQG 86
DB 61 LFGQHPDVFYLMPEPAHWMTKQSTAWMLHMAVFDLIRAVFLCDMSYFDAYMEGPRRQ 120
87 LFSQHPDVFYLMPEPAHWMTTISQSGSAATLHMAVFDLIRAVFLCDMSYFDAYMEGPRRQ 146
QY 121 SLSLFWENGRALCSAPACDIIIPQDEIIPRAHCRLLCSQPPFEVVEKACRSYSHVVLKEVR 180
DB 147 SLSLFWENGRALCSAPACDIIIPQDEIIPRAHCRLLCSQPPFEVVEKACRSYSHVVLKEVR 206
QY 181 FFNLSQSLYPLLKDPDSLNLHIVLVRDPAVFRSRETKGDLMDISRVNGQHEQKLKED 240

DB 207 FFNLSQSLYPLLKDPDSLNLHIVLVRDPAVFRSRETKGDLMDISRVNGQHEQKLKED 265
QY 241 QPYVMQVQCOSQLEIYK--TIQSLPKALQERYLLVREYEDLARAPVAQTSRMVYEFVGLGF 298
DB 266 PHLLIREVCKSHVRIAEATLKP-PFLGRVLRVRFEDLAREPLAEIRALYAFGTJTL 324
QY 299 LPHLQTVVHNTIRGKMGD--HAFHTNARDALNYSQAWRWSLPYKVKSRLOKACGDAMNL 356
DB 325 TPQLEAMHNTHTGSGIGKPIEAFHTSSRNARNYSQAWRHAFPTKILRVQEVCAQALQ 384
QY 357 LGYRHRSEQRNLLDLL-----STWVPE 383
DB 385 LGYRFPVSADQQRDLTDLVLPFGPDHFSWASPD 418
RESULT 7
ABB81557 standard; protein; 418 AA.
XX
XX ABB81557;
DT 05-SEP-2002 (first entry)
DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
OS Mus musculus.
XX
XX US2002061562-A1.
PD 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
XX (AKAM/) AKAMA T O.
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
DR New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
XX Example 5; Page 24-25; 69pp; English.
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyse sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratectomy. The present sequence
CC represents mouse intestinal N-acetylglucosamine-6-sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
XX Sequence 418 AA;
Query Match 50.4%; Score 1028; DB 5; Length 418;
Best Local Similarity 56.6%; Pred. No. 3.6e-99;
Matches 215; Conservative 43; Mismatches 112; Indels 10; Gaps 6;
QY 1 MLLPK--KMKLLFLVNSQMAILALFFHMYSHNIISSLSMKAKQPERMHVLSVSRSGSFV 58

XX	Rosen SD, Lee JK, Hemmerich S;
PI	
XX	
DR	WPI; 2001-139471/14.
DR	N-PDB; AAD02696.
XX	
PT	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT	diagnostic and therapeutic agent screening applications.
XX	
PS	Claim 3; Fig 2; 128pp; English.
XX	
CC	The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
CC	gene is found on chromosome 8E1. GSTR is a type 2 membrane protein useful
CC	for inhibiting a binding event between a selectin and a selectin ligand,
CC	which comprises contacting the selectin with a non-sulphated selectin
CC	ligand, GST and a small molecular agent that inhibits the sulphation
CC	activity of GST. GST is also useful in inhibiting a selectin mediated
CC	binding event. GST is useful in gene therapy to treat disorders such as
CC	acute or chronic inflammation, systemic lupus erythematosus (SLE),
CC	rheumatoid arthritis, polyarthritis nodosa, polymyositis,
CC	dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
CC	myaschemia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
CC	disease, adenailitis, hypoparathyroidism, pernicious anaemia,
CC	demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,
CC	myocarditis, regional enteritis, adult respiratory distress syndrome,
CC	infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial
CC	asthma, hypersensitivity, rheumatic fever and tissue rejection during
CC	transplantation
XX	
SQ	Sequence 395 AA;
	Query Match 49.9%; Score 1017.5; DB 4; Length 395;
	Best Local Similarity 56.4%; Pred. No. 4.3e-98;
	Matches 215; Conservative 473; Mismatches 112; Indels 11; Gaps 7;
QY	1 MLPLPK--RWKKLLFLVLSQMALIALFFHYSHNLISSLSMKAKQPERMHVLVLSWRSGGSFV 58
	: : :
Db	1 MELPFPSFTVMLSLIMTQTGLVF--LVSRQVPS--SPAGLGGRVHVHLVLSWRSGGSFV 56
	: : :

CY	S9	GOLFQCHEDVVFILMEPAWHVMIMFKUSTIAWMLHMAVDLLKAVFLCDDMSVIDAIMEPGR
Dd	S7	GOLFSQHDPDYLMPEPAWHVMDTLSSGSPALHMAVRDLIRSVFVLCDDMVDAYL-PWRR
Qy	119	ROSSIFOWENSRALCSAPACDIIPQBETIIPRAHCRLLCQQPFVEVEKACSYSHVVLKE
Dd	116	NISDLFWAVSRLCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACACSYSHVVLKE
Qy	179	VRFNNQLSLYPLLKDXPSLNLIHIVLRDPRVFRSRERTKGDLMDISRIWMGOHQELKK
Dd	176	VAFNQLQVLYPLSDPALNLRIHVLRDPRVLRSRERQTAKALARNDGVLTGTNGTWV-E
Qy	239	EQQPYVMQVTCQSGLIYK-TQTSIPKALOERYLLRVRYEDLARAPVACTSMYEYVCLE
Dd	235	ADPLRVNVNECRSHVRTAEAAHKPPFQDRYLVRVEDUARDPLTVIRELYFTGLG
Qy	298	FPLHLQTVWNHTIKGMKG--DHAPHNTARDALNVSQAWWSLPYKVSRLQKACGDAMN
Dd	295	LTPQLQTVHNTHITGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQ
Qy	356	LLGYRHSVRSEOEQRNLLDLL 376
Dd	355	LLGYRSVHSELEQRDLSDLL 375
 RESULT 12 AAU11275 ID AAU11275 standard; protein; 395 AA. XX AC XX AC XX DT XX DT XX XX DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.		

XX Mouse; beta1.3Gnt; beta1.3-N-acetylglucosaminyltransferase; MECA-79;
KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
KW antineoplastic; antiproliferative; antidiabetic; dermatological;
KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
XX Mus musculus.
XX WO200185177-A1.
XX 15-NOV-2001.
XX 10-MAY-2001; 2001WO-US015452.
XX 11-MAY-2000; 2000US-00569320.
XX (BURN-) BURNHAM INST.
XX Fukuda M, Yeh J, Hiraoka N;
XX WPI; 2002-075226/10.
XX N-PSDB; AAS16948.
XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
PT L-selectin sulfotransferase-2 that directs expression of L-selectin
PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
PT GlcNAc 6-sulfotransferase.
XX Claim 28; Fig 10; 98pp; English.
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated beta1.3-N-
CC acetylglucosaminyltransferase (beta1.3Gnt) or an active fragment, where
CC beta1.3Gnt directs expression of a MECA-79 antigen. The invention also
CC provides a method of treating or preventing an L-selectin-mediated
CC condition by reducing the expression or activity of a beta1.3Gnt that
CC directs expression of a MECA-79 antigen. This can be done by
CC administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1.3Gnt,
CC and/or a beta1.3Gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LSST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1.3Gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents mouse I-GlcNAc6ST
XX SQ Sequence 395 AA;
Query Match 49.9%; Score 1017.5; DB 5; Length 395;
Best Local Similarity 56.4%; Pred. No. 4.3e-98;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;
QY 1 MLPLK--KMKLLPLVSCMAILLAFHMYSHNLSLSMKQAQPERMVLVSSWRSSSV 58
Db 1 MRLPRFSTVMSLSLMTQTGLVLF---LVSRQVPS-SPAGLGERVHVVLSSWRSSSV 56
QY 59 GOLFGQHPDVFYLMPEAHVMMTPKQSTAMKLMHVRDLIRAVFLCDMSVFDAYMEPGPR 118
Db 57 GOLFSQHPDVFYLMPEAHVMMTDLSCSAPALHVAVEDLRSVFLCDMDVFDAYL-PWR 115
QY 119 QSSLSFQWNSRALCSAPADIIIPQDEIIIPRAHCRLLCSQOPPEVEVKACRSYSHVYLKE 178
Db 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVEVKPLCATRPFGLAQACSSYSHVYLKE 175
QY 179 VRFFNLQSLYLLKDPISLNLHIVLVRDPAVFRSRETRKGDLMIDSRVGMQGEQKLK 238

Db 176 VRFFNLQSLYLLKDPISLNLHIVLVRDPAVFRSRETRKALARDNGIVLTGNTWV-E 234
QY 239 EQOPYVWMOVICQSQLEIYK-TIOSLPKALQERYLLVRYEDLARAPVAQTSRMVFFVGL 297
Db 235 APRLEVVNVEVCRSHVIAEALHKPPPLQDRVRLVRYEDLARDPPLAVIRELYAFTGLG 294
QY 298 FPHLOTWVHNTTRKGWG--DHAFHTWARDALNVSOAWRWSLPYEKYSRLQKACGDAMN 355
Db 295 LTPQLQCTWHTNTHGSGPGARREAFKTTSRDALSVSQAWRHTLTPFAKRRVQELCGGALQ 354
QY 356 LLGYRHVRSEQRNLLDLL 376
Db 355 LLGYRSHSELEQRDLSLDLL 375
RESULT 13
ABB81555
ID ABB81555 standard; protein; 395 AA.
XX ABB81555;
XX AC ABB81555;
XX DT DT
XX 05-SEP-2002 (first entry)
XX Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX Homo sapiens.
XX Mus musculus.
XX Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 10 /label= Ala, Thr, Val
FT Misc-difference 13 /label= Ala, Val, Ser
FT Misc-difference 20 /label= Phe, Cys, Gly
FT Misc-difference 39 /label= Ala, Asp, Glu
FT Misc-difference 96 /label= Val, Met, Ile
FT Misc-difference 142 /label= Ala, Thr, Asn
FT Misc-difference 147 /label= Ala, Asp, Glu
FT Misc-difference 159 /label= Thr, Ser, Gly
FT Misc-difference 238 /label= Gly, His, Arg
FT Misc-difference 294 /label= Ser, Thr, Gly
FT Misc-difference 371 /label= Ala, Thr, Ser
FT Misc-difference 380 /label= Leu, Pro, Met
FT Misc-difference 382 /label= Gly, His, Ser
FT Misc-difference 384 /label= Thr, Ser, Lys
FT Misc-difference 390 /label= Ala, Glu
FT Misc-difference 391 /label= Ser, Lys
FT Misc-difference 392 /label= His, Gln
FT Misc-difference 394 /label= Arg, Glu
FT Misc-difference 395 /label= Asn, Ser
FT

XX US2002061562-A1.
XX 23-MAY-2002.
XX 09-AUG-2001; 2001US-00927602.
XX 11-AUG-2000; 2000US-00638211.
XX 11-AUG-2000; 2000US-0325773P.
XX (FUKU// FUKUDA M N.
XX (AKAM// AKAMA T O.
XX FukuDA MN, Akama TO;
XX WPI; 2002-507643/54.
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX Example 5; Fig 2A-B; 69pp; English.
XX The present invention describes human corneal N-acetylglucosamine-6-
XX sulfotransferase (I), which is able to catalyze sulfation of keratan
XX sulfate (KS). Also described is a method for monitoring the effect of
XX treatments for macular corneal dystrophy (MCD), and detecting
XX susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX ophthalmological activity. (II) can be used to treat or prevent macular
XX corneal dystrophy types I or II. (I) makes possible treatment of MCD
XX without requiring keratoplasty or keratectomy. The present sequence
XX represents a consensus N-acetylglucosamine-6-sulfotransferase which is
XX given in the exemplification of the present invention
XX
XX Sequence 395 AA;
XX
XX Query Match 49.7%; Score 1013.5; DB 5; Length 395;
XX Best Local Similarity 55.9%; Pred. No. 1.1e-97;
XX Matches 212; Conservative 45; Mismatches 101; Indels 21; Gaps 6;
XX
XX 2 LLPKXKLLFLVSVQMAILALFFHMYSHN1SLNKAQPERHVLVLSWRSQSGSFVQOL 61
XX 14 LLLAQTLXLLFLVSRP-----GPPSSPAGGEXRVRHVLVLSWRSQSGSFVQOL 59
XX 62 FQCHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEPGRQ 121
XX 60 FQCHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEPGRQ 118
XX 122 SLFQWENSRALCSAPACDIIPQDEIIPRAHCRLLCSQPFVEVEKACRSYSHVVLKEVRF 181
XX 119 DLFQWAVSRALCSPPACSAFPRGXISSEKVCPLCARQPFLLAREACRSYSHVVLKEVRF 178
XX 182 FNLQSLYPLKDPNLSNLHVLVRDPRVAFRSRETKGDLMDSDRIWQHEQKLKEDQ 241
XX 179 FNLQVLYPLSDPALNLRVHLVRDPRVAFRSRETKALARDNGIVLGTNGTWV-EADP 237
XX 242 PYVWMOVICQSOLEIYK--TIQSLPKALQERVLLVRYEDLARAPVAQTSRYVEFGLEFL 299
XX 238 XLRVVREVCESHVIRAEATLKP-PFFLEGRVLRVFEEDLAREPLAEIRALYAFGLXLT 296
XX 300 PHLQTVHNTGKNG--DHAFTNARDALNVSAWRSLPYEKVSLQKACGDAMNLL 357
XX 297 PQLAEIHNITGSGFGARREAFKTSRNALNVSAWRHALPFAKIRRVQELCAGALQL 356
XX 358 GYRHRVASEQQRNLLDL 376
XX 357 GYRPVYSEQRDLXLDIV 375
XX
XX RESULT 14
XX ID AAY72639
XX AAY72639 standard; protein; 390 AA.

AC AAY72639;
XX 02-MAY-2001 (first entry)
XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarteritis nodosa; polymyositis; systemic scleritis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1.
XX Homo sapiens.
XX WO200106015-A1.
XX 25-JAN-2001.
XX 19-JUL-2000; 2000WO-US019741.
XX 20-JUL-1999; 99US-0144694P.
XX 13-JUN-2000; 2000US-00593828.
XX (REGC) UNIV CALIFORNIA.
XX Rosen SB, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
XX N-PSDB; AAD02697, AAD02698, AAD02699.
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications.
XX Claim 3; Fig 1; 128pp; English.
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
XX membrane protein useful for inhibiting a binding event between a selectin
XX and a selectin ligand, which comprises contacting the selectin with a non
XX -sulphated selectin ligand, GST and a small molecular agent that inhibits
XX the sulphation activity of GST. GST is also useful in inhibiting a
XX selectin mediated binding event. GST is useful in gene therapy to treat
XX disorders such as acute or chronic inflammation, systemic lupus
XX erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
XX polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation. Note: the present sequence is also shown in
XX sequence listing (page no: 56) but lacks four nucleotides at its 3' end.
XX Sequence 390 AA;
XX
XX Query Match 49.5%; Score 1008; DB 4; Length 390;
XX Best Local Similarity 52.3%; Pred. No. 4.2e-97;
XX Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
XX
XX 1 MLLPKXKLLFLVSVQMAILALFFHMYSHN1SLNKAQPERHVLVLSWRSQSGSFVQOL 60
XX 14 LLLAQTLXLLFLVSRP-----GPPSSPAGGEXRVRHVLVLSWRSQSGSFVQOL 59
XX 61 LFCQHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
XX 60 LFSQHPDVFLMEPAHVMWTFKQSTAWMLHVAVRDLIRAVFLCDMSVFDAYMEPGRQ 118

QY 121 SSLFOWNSRALCSAPACDIIIPQDEIIIPRAHCRLLCSQOPPEVVEKACRSYSHVVLKEVR 180
DB 119 SAFFWATSRALCSPACSAFPGTISKQDVCKTLCTQPSFLAREACRSYSHVVLKEVR 178
QY 181 FPNLSQVPLLLKDPKSLNLIHIVLVRDPAVFRSRERTKGDLMDSRIWNGHQEKKKED 240
DB 179 FPNLSQVPLLLKDPKSLNLIHIVLVRDPAVFRSRERTKGDLMDSRIWNGHQEKKKED 237
QY 241 QPYVYVMOVICQSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAOTSRYMYEFVGLGF 298
DB 238 PHLRLEIVCRSHVRIAEANLTKP--PPFLRGYRLVRVEDLAREPLAIRALYFTGLTL 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVQAWRWSLPYKVSRLQKACGDAMNL 356
DB 297 TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVQAWRHALPFTKILRVQEVCAQALQ 356
QY 357 LGYRVHRSQEQORNLLDLL-----STWTVPE 383
DB 357 LGYRVHRSQEQORNLLDLL-----STWTVPE 390

RESULT 15
ID ABB81556
XX ABB81556 standard; protein; 390 AA.
AC ABB81556;
XX
DT 05-SEP-2002 (first entry)
XX Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
DE Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc-6ST;
KW cornelal; sulfation; Keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX Homo sapiens.
XX US2002061562-A1.
XX 23-MAY-2002.
XX 09-AUG-2001; 2001US-00927602.
XX 11-AUG-2000; 2000US-00638211.
XX 11-AUG-2000; 2000US-0325773P.
XX (FUKU/) FUKUDA M N.
PA (AKAM/) AKAWA T O.
XX
XX Fukuda MN, Akama TO;
XX WPI; 2002-507643/54.
XX
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX
XX Example 5; Fig 2A-B; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
XX sulfotransferase (I), which is able to catalyse sulfation of keratan
XX sulfate (KS). Also described is a method for monitoring the effect of
XX treatments for macular corneal dystrophy (MCD), and detecting
XX susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX ophthalmological activity. (I) can be used to treat or prevent macular
XX corneal dystrophy types I or II. (I) makes possible treatment of MCD
XX without requiring keratinoplasty or keratectomy. The present sequence
XX represents human intestinal N-acetylglucosamine-6-sulfotransferase,
XX which is given in comparison with (I) in the exemplification of the
XX present invention
XX Sequence 390 AA;

Query Match 49.5%; Score 1008; DB 5; Length 390;
Best Local Similarity 52.3%; Pred.No. 4.2e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 1 MLLPKKKKLLFLVSSQMAILALPFFHMYSHNIISSLSMKQAQPERMHVILVSSWSGSGSFVQ 60
DB 14 LLLAQTTCLLLFTIIRP-----CPSSPAGGEDRVHVLVSSWSGSGSFVQ 59
QY 61 LFOQHPDVFLMEPAWHVMTTKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRKQ 120
DB 60 LFSQHPDVFLMEPAWHVMTTKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRKQ 118
QY 121 SSLFOWNSRALCSAPACDIIIPQDEIIIPRAHCRLLCSQOPPEVVEKACRSYSHVVLKEVR 180
DB 119 SAFFWATSRALCSPACSAFPGTISKQDVCKTLCTQPSFLAREACRSYSHVVLKEVR 178
QY 181 FPNLSQVPLLLKDPKSLNLIHIVLVRDPAVFRSRERTKGDLMDSRIWNGHQEKKKED 240
DB 179 FPNLSQVPLLLKDPKSLNLIHIVLVRDPAVFRSRERTKGDLMDSRIWNGHQEKKKED 237
QY 241 QPYVYVMOVICQSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAOTSRYMYEFVGLGF 298
DB 238 PHLRLEIVCRSHVRIAEANLTKP--PPFLRGYRLVRVEDLAREPLAIRALYFTGLTL 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVQAWRWSLPYKVSRLQKACGDAMNL 356
DB 297 TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVQAWRHALPFTKILRVQEVCAQALQ 356
QY 357 LGYRVHRSQEQORNLLDLL-----STWTVPE 383
DB 357 LGYRVHRSQEQORNLLDLL-----STWTVPE 390

Search completed: May 6, 2004, 10:49:57
Job time : 60 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:51:28 ; Search time 48 Seconds
(without alignments)
2232.099 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKMKLLFLVQMAIL.....EORNLLDLLSTWTVEQIH 386

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/prodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/1/pubaa/US09D_PUBCOMB.pep.*
- 13: /cgn2_6/prodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/1/pubaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/prodata/1/pubaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/prodata/1/pubaa/US10F_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2038	100.0	386	9 US-09-816-825-2	Sequence 2, Appli
2	2038	100.0	386	13 US-10-007-262-1	Sequence 1, Appli
3	1936	95.0	386	14 US-10-427-631-11	Sequence 11, Appl
4	1028	50.4	418	9 US-09-927-602-5	Sequence 5, Appli
5	1019	50.0	395	9 US-09-927-602-2	Sequence 2, Appli
6	1019	50.0	395	12 US-10-258-080-5	Sequence 5, Appli
7	1013.5	49.7	395	9 US-09-927-602-3	Sequence 3, Appli
8	1008	49.5	390	9 US-09-927-602-4	Sequence 4, Appli
9	821	40.3	171	9 US-09-927-602-8	Sequence 8, Appli
10	598.5	29.4	483	14 US-10-212-933-2	Sequence 2, Appli
11	587.5	28.8	484	14 US-10-212-933-4	Sequence 4, Appli
12	587.5	28.8	531	9 US-09-833-790-255	Sequence 255, Appl
13	549	26.9	411	12 US-10-211-462-97	Sequence 97, Appl
14	549	26.9	411	14 US-10-021-660-128	Sequence 128, Appl
15	519.5	25.5	169	9 US-09-927-602-6	Sequence 6, Appli

16	511.5	25.1	169	9	US-09-927-602-7	Sequence 7, Appli
17	510.5	24.8	481	12	US-10-087-192-123	Sequence 123, App
18	488	23.9	479	12	US-10-087-192-126	Sequence 126, App
19	336	16.5	169	9	US-09-927-602-9	Sequence 9, Appli
20	335.5	16.5	179	9	US-09-927-602-10	Sequence 10, Appli
21	293	14.4	174	9	US-09-927-602-11	Sequence 11, Appli
22	116	5.7	387	14	US-10-126-279-21	Sequence 21, Appl
23	116	5.7	387	14	US-10-286-608-21	Sequence 21, Appl
24	100.5	4.9	668	10	US-09-811-469-6	Sequence 6, Appli
25	100.5	4.9	668	14	US-10-370-659-6	Sequence 39283, A
26	100.5	4.9	807	12	US-10-425-114-39283	Sequence 854, App
27	100	4.9	593	12	US-10-225-066A-854	Sequence 56, Appl
28	100	4.9	593	15	US-10-374-780A-330	Sequence 330, App
29	100	4.9	593	15	US-10-374-780A-330	Sequence 149380,
30	98	4.8	629	16	US-10-380-727-16	Sequence 16, Appl
31	96.5	4.7	791	16	US-10-380-727-16	Sequence 38, Appl
32	95.5	4.7	568	12	US-10-042-865-38	Sequence 132, App
33	95.5	4.7	568	15	US-10-262-445-132	Sequence 380, App
34	93	4.6	348	15	US-10-259-194A-380	Sequence 11, Appl
35	92	4.5	359	14	US-10-411-976-11	Sequence 1075, Ap
36	91	4.5	1847	15	US-10-369-493-1075	Sequence 2, Appli
37	90	4.4	393	14	US-10-126-279-2	Sequence 2, Appli
38	90	4.4	393	14	US-10-286-606-2	Sequence 23144, A
39	89.5	4.4	4273	15	US-10-369-493-23144	Sequence 20, Appl
40	89	4.4	320	9	US-09-854-122-20	Sequence 76479, A
41	89	4.4	1140	12	US-10-282-122A-76479	Sequence 15, Appl
42	89	4.4	1968	14	US-10-223-070-15	Sequence 62635, A
43	88	4.3	746	12	US-10-425-114-62635	Sequence 1903, Ap
44	88	4.3	2209	9	US-09-902-941-1903	Sequence 1903, Ap
45	88	4.3	2209	9	US-09-849-626-1903	

ALIGNMENTS

RESULT 1
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Query Match	100.0%;	Score	2038;	DB	9;	Length	386;
Best Local Similarity	100.0%;	Pred. No.	4.6e-202;				
Matches	386;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MLLPKMKLLFLVQMAILALFFHMYSHN1SSLSMKAQPERMHVLSWSRSGSFVQ	60				
Db	1	MLLPKMKLLFLVQMAILALFFHMYSHN1SSLSMKAQPERMHVLSWSRSGSFVQ	60				
QY	61	LFQGPDPVFLMEPAHWVMTFKQSTAWMLHNAVRLIRAVFLCDMSVFDYMEPGRQ	120				
Db	61	LFQGPDPVFLMEPAHWVMTFKQSTAWMLHNAVRLIRAVFLCDMSVFDYMEPGRQ	120				
QY	121	SSLFOWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOPEVVEKACRSYSHVVLKEVR	180				
Db	121	SSLFOWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQOPEVVEKACRSYSHVVLKEVR	180				

QY 181 FFLQSLYPLKDPKPSLNHLVHLVRDPRVFRSRETKGDLMDSDRIVMGQHEQKLKED 240
Db 181 FFLQSLYPLKDPKPSLNHLVHLVRDPRVFRSRETKGDLMDSDRIVMGQHEQKLKED 240
QY 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
Db 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
QY 301 HLQTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
Db 301 HLQTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
QY 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

RESULT 2

US-10-007-262-1
; Sequence 1, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: COSLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yaida
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-007-262-1

Query Match 100.0%; Score 2038; DB 13; Length 386;
Best Local Similarity 100.0%; Pred. No. 4.6e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVILVLSWRSFGVQ 60
Db 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVILVLSWRSFGVQ 60
QY 61 LFGQHPDVFLMEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDATMEGPRRQ 120
Db 61 LFGQHPDVFLMEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDATMEGPRRQ 120
QY 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
Db 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
QY 181 FFLQSLYPLKDPKPSLNHLVHLVRDPRVFRSRETKGDLMDSDRIVMGQHEQKLKED 240
Db 181 FFLQSLYPLKDPKPSLNHLVHLVRDPRVFRSRETKGDLMDSDRIVMGQHEQKLKED 240
QY 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
Db 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
QY 301 HLQTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
Db 301 HLQTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
QY 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

RESULT 4

US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:

RESULT 3

US-10-427-631-11
; Sequence 11, Application US/10427631
; Publication No. US20030175923A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
; APPLICANT: COSLEY, Neil C.; GUEGLER, Karl J.;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
; APPLICANT: AZIMZAI, Yaida
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592-1 DIV
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 09/786,240
; PRIOR APPLICATION NUMBER: PCT/US99/20989
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: US 60/172,220
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US 60/155,248
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/133,642
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-427-631-11

Query Match 95.0%; Score 1936; DB 14; Length 386;
Best Local Similarity 95.6%; Pred. No. 1.7e-191;
Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVILVLSWRSFGVQ 60
Db 1 MLLPKKMKLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVILVLSWRSFGVQ 60
QY 61 LFGQHPDVFLMEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDATMEGPRRQ 120
Db 61 LFGQHPDVFLMEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDATMEGPRRQ 120
QY 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
Db 121 SSFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPFEVVEKACRSYSHVVLKEVR 180
QY 181 FFLQSLYPLKDPKPSLNHLVHLVRDPRVFRSRETKGDLMDSDRIVMGQHEQKLKED 240
Db 181 FFLQSLYPLKDPKPSLNHLVHLVRDPRVFRSRETKGDLMDSDRIVMGQHEQKLKED 240
QY 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
Db 241 QPYVWQVICSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVACTSRMYEFVGLFPL 300
QY 301 HLQTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
Db 301 HLQTWVHNITRGKMGDGHAFHTNARDALNVQAWRWSLPYEKVSRLOKACGDMNLLGYR 360
QY 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPEQIH 386

APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US 09/927,602
PRIOR FILING DATE: 2001-08-09
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 418
TYPE: PRT
ORGANISM: Mus musculus
US-09-927-602-5

Query Match 50.4%; Score 1028; DB 9; Length 418;

Best Local Similarity 56.6%; Pred. No. 2e-97;
Matches 215; Conservative 43; Mismatches 112; Indels 10; Gaps 6;

QY 1 MLLPK--KMKLLFLVSQMAILALPFHMYSHNLSMKQAQPERMHVLSVSRSSGFV 58
DB 25 MLLPSSVTMLSLMVQGLVFP---LVSROVPS-SPAGLGERVHVLSVSRSSGFV 80
QY 59 GQFGQHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEQPR 118
DB 81 GQFSGHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYL-PWR 139
QY 119 ROSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLSCQPFVVEKACRSYSHVLE 178
DB 140 NISDLFQWAVSRALCSAPACDIIPQDEIIPRAHCRLLSCQPFVVEKACRSYSHVLE 199
QY 179 VRFNLQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLK 238
DB 200 VRFNLQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLK 258
QY 239 EDPQVYVMOVICOSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 298
DB 259 ADPLRVNVECRSHVRIAEATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYFTGL 318
QY 299 LPHQTVWNTIRGKMG--DHAFTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 356
DB 319 TPQLQTVWNTIRGKMG--DHAFTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 378
QY 357 LGYRVSRQEQRLNLLDLL-----STWTVPEQIH 386
DB 379 LGYRVSRQEQRLNLLDLL-----STWTVPEQIH 398

RESULT 5

US-09-927-602-2
Sequence 2, Application US/09927602
Patent No. US2002061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 395
TYPE: PRT
ORGANISM: Homo Sapien
US-09-927-602-2

Query Match 50.0%; Score 1019; DB 9; Length 395;

Best Local Similarity 54.5%; Pred. No. 1.6e-96;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;
QY 2 LLPKMKLLFLVSQMAILALPFHMYSHNLSMKQAQPERMHVLSVSRSSGFVQGL 61
DB 14 LLLAQTFLLFLVSRP-----GPFSSPAGGEARVHVLSVSRSSGFVQGL 59
QY 62 FGQHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEQPRROS 121
DB 60 FNQHPDVFYLMPEAHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYL-PWRNLS 118
QY 122 SLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLSCQPFVVEKACRSYSHVLEKVRP 181
DB 119 DLQWAVSRALCSAPACDIIPQDEIIPRAHCRLLSCQPFVVEKACRSYSHVLEKVRP 178
QY 182 FNLQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLKDEQ 241
DB 179 FNLQSLYPLKPSLNLHVLVDRPRAVFRSRTKGLMIDSRIVMGQHQKLLKDEQ 237
QY 242 PYVMOVICOSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 299
DB 238 GLRVVRECRSHVRIAEATLKP-PPFLRGYRLVRFEDLAREPLAEIRALYFTGLSLT 296
QY 300 PHQTVWNTIRGKMG--DHAFTNARDALNVSOAWRWSLPYKVSRLQKACGDAMNL 357
DB 297 POLEAMHNTHTGSGPGARREAFKTSRNLNVSOAWRWSLPYKVSRLQKACGDAMNL 356
QY 358 GYRVSRQEQRLNLLDLL-----STWTVPEQIH 386
DB 357 GYRVSRQEQRLNLLDLL-----STWTVPEQIH 392

RESULT 6

US-10-258-080-5
Sequence 5, Application US/10258080
Publication No. US20040029125A1
GENERAL INFORMATION:
APPLICANT: Incyte Genomics, Inc.
APPLICANT: POLICKY, Jennifer L.
APPLICANT: HAPALIA, April J.A.
APPLICANT: BURFORD, Neil
APPLICANT: RING, Huijun Z.
APPLICANT: LAL, Preeti G.
APPLICANT: TRIBOULEY, Catherine M.
APPLICANT: YAO, Monique G.
APPLICANT: YUE, Henry
APPLICANT: TANG, Y. Tom
APPLICANT: ARVIZU, Chandra S.
APPLICANT: DAS, Debopriya
APPLICANT: SANJANWALA, Madhusudan M.
APPLICANT: GANDHI, Ameena R.
APPLICANT: REDDY, Roopa M.
APPLICANT: KHAN, Farrah A.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: RAKUNAR, Jayalaxmi
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: AU-YOUNG, Janice K.
TITLE OF INVENTION: DRUG METABOLIZING ENZYMES
FILE REFERENCE: PI-0070 USN
CURRENT APPLICATION NUMBER: US/10/258,080
CURRENT FILING DATE: 2002-10-15
PRIOR APPLICATION NUMBER: US 60/203,509
PRIOR FILING DATE: 2000-05-11
PRIOR APPLICATION NUMBER: US 60/202,234
PRIOR FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/200,185
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/198,403
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US01/11869
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: US 60/197,590
PRIOR FILING DATE: 2000-04-13

```
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040029125A1 7472777CD1
US-10-258-080-5

Query Match          50.0%; Score 1019; DB 12; Length 395;
Best Local Similarity 54.5%; Pred. No. 1.6e-96;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY  2 LLPKMKLLFLVSOAILALFFHMYSHNISLSMKAQPERMHVILVLSWSSGSGSFVQQL 61
DB  14 LLLAQTLFLVSRP-----GSSSPAGGEARVHVILVLSWSSGSGSFVQQL 59
QY  62 FQCHPDVFLMEPAWVHMTFKQSTAMMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 121
DB  60 FQCHPDVFLMEPAWVHMTFKQSTAMMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 118
QY  122 SLFQWNSRALCSAPACDIIPDEIIPRAHCELLCSQPFVEVEKACRSYSHVILKEVRF 181
DB  119 DLFWAVSRALCSPPACSAFPRGXISSEKVCPLCARQPFKLAREACRSYSHVILKEVRF 178
QY  182 FNLQSLYPLLKDPSPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 241
DB  179 FNLQVLYPLSDPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 237
QY  242 PYYMQVICOSQLEIYK--TIQSLKALQERYLLVRYEDLARAPVAQTSRMYEFVGLLEFL 299
DB  238 GLRVVREVCRSVRVIAEAATLKP-PFFLRGRYLRVRFEDLAREPLAEIRALYAFGLSLT 296
QY  300 PHLOTWVHNITRGKMG--DHAFHTNARDALNVSAWRWSLPYEKVSRLQKACGDMNLL 357
DB  297 PQLEAWIHNITGSGPGAREAFKTSRNALNVSAWRHALPFAKIRVQELCAGALQLL 356
QY  358 GYRHRVSEQORNLDDL 376
DB  357 GYRVPVSEQORNLDDL 375

RESULT 8
US-09-927-602-4
; Sequence 4, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-4

Query Match          49.5%; Score 1008; DB 9; Length 390;
Best Local Similarity 52.3%; Pred. No. 2.1e-95;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY  1 MLLPKMKLLFLVSOAILALFFHMYSHNISLSMKAQPERMHVILVLSWSSGSGSFVQ 60
DB  14 LLLAQTLFLVSRP-----GSSSPAGGEDRVHVILVLSWSSGSGSFVQ 59
QY  61 FQCHPDVFLMEPAWVHMTFKQSTAMMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 120
DB  60 FQCHPDVFLMEPAWVHMTFKQSTAMMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 118
QY  121 SSLFQWNSRALCSAPACDIIPDEIIPRAHCELLCSQPFVEVEKACRSYSHVILKEVRF 180
DB  119 SAFFNWATSRALCSPPACSAFPRGTTISKQDVCKTLCTROPFLAREACRSYSHVILKEVRF 178
QY  181 FNLQSLYPLLKDPSPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 240
DB  179 FNLQVLYPLSDPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 237

; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040029125A1 7472777CD1
US-10-258-080-5

Query Match          50.0%; Score 1019; DB 12; Length 395;
Best Local Similarity 54.5%; Pred. No. 1.6e-96;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY  2 LLPKMKLLFLVSOAILALFFHMYSHNISLSMKAQPERMHVILVLSWSSGSGSFVQQL 61
DB  14 LLLAQTLFLVSRP-----GSSSPAGGEARVHVILVLSWSSGSGSFVQQL 59
QY  62 FQCHPDVFLMEPAWVHMTFKQSTAMMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 121
DB  60 FQCHPDVFLMEPAWVHMTFKQSTAMMLHMAVRDLIRAVFLCDMSVDFDAYMEPGRQ 118
QY  122 SLFQWNSRALCSAPACDIIPDEIIPRAHCELLCSQPFVEVEKACRSYSHVILKEVRF 181
DB  119 DLFWAVSRALCSPPACSAFPRGXISSEKVCPLCARQPFKLAREACRSYSHVILKEVRF 178
QY  182 FNLQSLYPLLKDPSPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 241
DB  179 FNLQVLYPLSDPALNLHIVLRDPRAVFRSRETKGDLMDSDRVNMGQHEQKLKEDQ 237
QY  242 PYYMQVICOSQLEIYK--TIQSLKALQERYLLVRYEDLARAPVAQTSRMYEFVGLLEFL 299
DB  238 GLRVVREVCRSVRVIAEAATLKP-PFFLRGRYLRVRFEDLAREPLAEIRALYAFGLSLT 296
QY  300 PHLOTWVHNITRGKMG--DHAFHTNARDALNVSAWRWSLPYEKVSRLQKACGDMNLL 357
DB  297 PQLEAWIHNITGSGPGAREAFKTSRNALNVSAWRHALPFAKIRVQELCAGALQLL 356
QY  358 GYRHRVSEQORNLDDL 376
DB  357 GYRVPVSEQORNLDDL 375

RESULT 7
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; NAME/KEY: VARIANT
; LOCATION: (1)...(395)
; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match          49.7%; Score 1013.5; DB 9; Length 395;
Best Local Similarity 55.9%; Pred. No. 5.9e-96;
```



```

; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-212-933-4

Query Match      28.8%; Score 587.5; DB 14; Length 484;
Best Local Similarity 36.0%; Pred. No. 1e-51;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps
QY      41  ERMHVLVLSWSGSGSFVQQLFGQHPDVYFLMEPAHVMWMTFKQSTAWMLHMAVRDLIRA 100
Db      117  KRHWMTVFTWRSKSGSFGELFNQNPVEFFLYEPVHWQKLYPGDAVSLQGAARDMLSA 176
QY      101  VFLCDMSVFDAYMEFCPPRQS-----SLFPQWNSRALCSAPACDIIIPQDEIIPRAHCRLL- 155
Db      177  LYRCDLSVFQLYSPAGSGGRNLTLTGIFGAATNKVCGSSPLCPAY-RKEVGVGLVDDRCK 235
QY      156  -CSQQPFEVWEKACHSYSHVVLKVPFNLOSILPKDPSLNLIHVLVRPRAVFRSR 214
Db      236  KCPQRLARFEECKRYTLVKGVRVEDVAFLPLRDPAIDLKVIHLVRDPRAVASSR 295
QY      215  ERTKGLMDISIRM-----CQHEOKLKKE---DQPYV---VMOVICOS 252
Db      296  IRSRGLTRESLVVRSRDRPRAHRMPFLAAGHKLGAKEGVGPGADYHALGNMEVICS 355
QY      253  QLEIVKTIOSLPKALQERYLLVRVEDLARAPVQTSRMVEFVGLFPLHLOTWVNIIRG 312
Db      356  MAKTLQATALQPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALNMTSG 415
QY      313  KMGGDHAFHTNARDALNVSQAWRSLLPYEKVSRLLQKACGDAMNLGVYRHSRSEQORNL 372
Db      416  SGSSSKPFFVVSARNATQANAWRTALFQQIKQVEEFCYQPMVLGVYRNSPEEVKDL 475
QY      373  LDLL 376
Db      476  KTL 479

RESULT 12
US-09-833-790-255
; Sequence 255, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liguin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121 512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 255
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-790-255

Query Match      28.8%; Score 587.5; DB 9; Length 531;
Best Local Similarity 36.0%; Pred. No. 1.2e-51;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps
QY      41  ERMHVLVLSWSGSGSFVQQLFGQHPDVYFLMEPAHVMWMTFKQSTAWMLHMAVRDLIRA 100
Db      164  KRHWMTVFTWRSKSGSFGELFNQNPVEFFLYEPVHWQKLYPGDAVSLQGAARDMLSA 223

```

```
QY 157 SQPFVWVKACRSYSHVVLKEVRFENLQSLYPLKDPNSLNHIVHLVRDPAVRSRER 216
Db 186 GLNLNLTVAEACRERSHVAIKTVRPEVNDLRAVEDPRLNKLVIQLVRDPGILASRSE 245
QY 217 TKGDLMSDRIVMGQHEQKLKEDOPYV---MQVICSQLEIYKTIQSLPKALQERYL 272
Db 246 TFRDYRLNRLWYGTGR-----KPYNLDVTQLTTCDFNSVSTGLMRPPWLKGYM 298
QY 273 LVRYEDLARAPVAQTSRMVYFVGLBFLPHLOTWVHNITRG-KGMGDHAFHTNARDALNVS 331
Db 299 LVRYEDLARPNMKTEETIYGLGIPLDSDHVARWIONNTRGDDPTLGKHKYGTVRNSAATA 357
QY 332 QAWNSLPYKVSRLQKACGDAMNLLGYHRVSEGEORNLILLDL 376
Db 358 EKWRFRLSYDIVAFNAQCOQVLAQGLGYKIRASEBELKNPVS LV 402

RESULT 14
US-10-021-660-128
; Sequence 128, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynnne, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-000710US
; CURRENT APPLICATION NUMBER: US/10/021,660
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-128

Query Match 26.9%; Score 549; DB 14; Length 411;
Best Local Similarity 34.3%; Pred. No. 7.8e-48;
Matches 139; Conservative 64; Mismatches 160; Indels 42; Gaps 10;

QY 6 KMKLLJLLFLYS---QWAILALFPHMYSHNLISSLMKAQPERM-----HVLV 47
Db 6 KAVLLALASIAIQYTAINTFTAKSFHTCPGLAEAGLAERLCEESPTTFAYNLKRKTHILI 65
QY 48 LSSWRSGSFVGLQFGQHPDVFYLMPEAWHVMWT----FKQSTA----WMLHMAVRDLIR 99
Db 66 LATTRSGSFVGLQFNQHLDDVLYLFEPLHVQNTLIPFTQKSPADRRVMLGASRDLLR 125
QY 100 AVFLCDMSVFDAYMEPGPRQSS--LFWENSRALCSAPACDII-PQDEIIPRAHCRLLC 156
Db 126 SLYDCDLYLENIYKIPPPVNNHTTDRIFRAGASRLCSRPVCDPPGPGADLVLEEGDCVKEC 185
QY 157 SQPFVWVKACRSYSHVVLKEVRFENLQSLYPLKDPNSLNHIVHLVRDPAVRSRER 216
Db 186 GLNLNLTVAEACRERSHVAIKTVRPEVNDLRAVEDPRLNKLVIQLVRDPGILASRSE 245
QY 217 TKGDLMSDRIVMGQHEQKLKEDOPYV---MQVICSQLEIYKTIQSLPKALQERYL 272
Db 246 TFRDYRLNRLWYGTGR-----KPYNLDVTQLTTCDFNSVSTGLMRPPWLKGYM 298
QY 273 LVRYEDLARAPVAQTSRMVYFVGLBFLPHLOTWVHNITRG-KGMGDHAFHTNARDALNVS 331
Db 299 LVRYEDLARPNMKTEETIYGLGIPLDSDHVARWIONNTRGDDPTLGKHKYGTVRNSAATA 357
QY 332 QAWNSLPYKVSRLQKACGDAMNLLGYHRVSEGEORNLILLDL 376
```

```
Db 358 EKWRFRLSYDIVAFNAQCOQVLAQGLGYKIRASEBELKNPVS LV 402

RESULT 15
US-09-927-602-6
; Sequence 6, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Tomoya O.
; APPLICANT: Akama, Tomoya N.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-927-602-6

Query Match 25.5%; Score 519.5; DB 9; Length 169;
Best Local Similarity 41.1%; Pred. No. 2.5e-45;
Matches 113; Conservative 15; Mismatches 34; Indels 113; Gaps 3;

QY 42 RMHVLVLSWRSGSFVGLQFGQHPDVFYLMPEAWHVMWTFKQSTAWMLHMAVRDLIRAV 101
Db 8 RVHVLVLSWRSGSFVGLQFGQHPDVFYLMPEAWHVMWT----- 47
QY 102 FLCDSVFDAYMEPGPRROSSLPQWENSALCSAPACDIIPODEIIPRAHCRLLCSQQPF 161
Db 48 ----- 47
QY 162 EVVEKACRSYSHVVLKEVRFENLQSLYPLKDPNSLNHIVHLVRDPAVRSRERTKGD 221
Db 48 --LSEACRSYSHVVLKEVRFENLQSLYPLSDPALNRIVHLVRDPAVRSRERTKAKAL 105
QY 222 MIDSRIVMGQHEQKLKEDOPYVVMQVICSQLEIYKTIQSLPKALQERYLLVRYEDLAR 281
Db 106 ARDNEAA-----TLKP-PPFLRGYRLVRFEDLAR 134
QY 282 APVAQTSRMVYFVGLBFLPHLOTWVHNITRGKMG 316
Db 135 EPLAETRALYFTGLSLTPOLEAWIHNIHTGSGFG 169

Search completed: May 6, 2004, 10:57:20
Job time : 49 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:48:57 ; Search time 22 Seconds
(without alignments)
905.801 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLLPKMKLLFLVSQMAIL.....EQRNLLDLLSTWTVPQI 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/aaa/5A COMB.pdp:*
2: /cgn2_6/prodata/2/aaa/5B COMB.pdp:*
3: /cgn2_6/prodata/2/aaa/6A COMB.pdp:*
4: /cgn2_6/prodata/2/aaa/6B COMB.pdp:*
5: /cgn2_6/prodata/2/aaa/PCTUS COMB.pdp:*
6: /cgn2_6/prodata/2/aaa/backfiles1.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2038	100.0	386	3	US-09-045-284A-2
2	2038	100.0	386	4	US-09-190-911-1
3	1936	95.0	386	4	US-09-786-240-11
4	598.5	29.4	483	3	US-09-263-023-2
5	598.5	29.4	483	4	US-09-471-867-2
6	587.5	28.8	484	3	US-09-263-023-4
7	587.5	28.8	484	4	US-09-471-867-4
8	549	26.9	411	4	US-09-015-188-2
9	500.5	24.6	458	2	US-08-655-878-2
10	482	23.7	479	2	US-08-899-514-2
11	100.5	4.9	668	4	US-09-811-469-6
12	94.5	4.6	848	4	US-09-540-824-27
13	92	4.5	359	3	US-09-150-133-11
14	92	4.5	359	3	US-09-150-141-11
15	92	4.5	359	3	US-09-374-493-11
16	92	4.5	359	3	US-09-374-824-11
17	92	4.5	359	3	US-09-374-492-11
18	92	4.5	359	4	US-09-785-343-11
19	87.5	4.3	380	3	US-09-150-133-9
20	87.5	4.3	380	3	US-09-150-141-9
21	87.5	4.3	380	3	US-09-374-493-9
22	87.5	4.3	380	3	US-09-374-824-9
23	87.5	4.3	380	4	US-09-374-492-9
24	87.5	4.3	380	4	US-09-785-343-9
25	86.5	4.2	831	1	US-08-073-384C-5
26	86.5	4.2	831	1	US-08-254-359A-5
27	86.5	4.2	831	1	US-08-483-043-5

28	86.5	4.2	831	1	US-08-481-238-5	Sequence 5, Appli
29	86.5	4.2	831	2	US-08-471-066B-5	Sequence 5, Appli
30	86.5	4.2	831	2	US-08-484-956-5	Sequence 5, Appli
31	86.5	4.2	831	2	US-08-757-653-5	Sequence 5, Appli
32	86.5	4.2	831	2	US-08-599-491-5	Sequence 5, Appli
33	86.5	4.2	831	2	US-08-756-386-5	Sequence 5, Appli
34	86.5	4.2	831	2	US-08-823-516-5	Sequence 5, Appli
35	86.5	4.2	831	3	US-08-882-853A-5	Sequence 5, Appli
36	86.5	4.2	831	3	US-08-759-038-5	Sequence 5, Appli
37	86.5	4.2	831	3	US-08-758-314-5	Sequence 5, Appli
38	86.5	4.2	831	4	US-09-350-309-5	Sequence 5, Appli
39	86.5	4.2	831	4	US-08-520-946-5	Sequence 5, Appli
40	86.5	4.2	831	4	US-09-684-938-5	Sequence 5, Appli
41	86.5	4.2	831	4	US-09-308-825A-5	Sequence 5, Appli
42	86.5	4.2	831	4	US-09-758-282B-5	Sequence 5, Appli
43	86.5	4.2	831	4	US-09-655-378A-5	Sequence 5, Appli
44	84.5	4.1	2549	4	US-08-265-967C-1	Sequence 1, Appli
45	84.5	4.1	2549	4	US-08-305-790B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 100.0%; Score 2038; DB 3; Length 386;					
Best Local Similarity 100.0%; Pred. No. 38-212;					
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Oy	1	MLLPKMKLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVLVLSWSRSGSFVQG	60		
Db	1	MLLPKMKLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVLVLSWSRSGSFVQG	60		
Oy	61	LFQGHDPVFLMSPAHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGRRQ	120		
Db	61	LFQGHDPVFLMSPAHVWMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGRRQ	120		
Oy	121	SSLFQWNSRSLCSAPACDIIPODEIIPRAHCHLLCSQPFVEVVEKACRSYSHVLEKVR	180		
Db	121	SSLFQWNSRSLCSAPACDIIPODEIIPRAHCHLLCSQPFVEVVEKACRSYSHVLEKVR	180		
Oy	181	FFNLQSLYPLKDPSLNLHIVLVRDPRAVFRSRRRTKGDLMIDSRVIMGQHEQKLKED	240		
Db	181	FFNLQSLYPLKDPSLNLHIVLVRDPRAVFRSRRRTKGDLMIDSRVIMGQHEQKLKED	240		
Oy	241	QPYVWMOVICQSQLEIYKTIQSLPKALQERYLLVRVEDLARAPVAQTSRMRYEFGLEFLP	300		
Db	241	QPYVWMOVICQSQLEIYKTIQSLPKALQERYLLVRVEDLARAPVAQTSRMRYEFGLEFLP	300		
Oy	301	HLQTVHNIITRGKMGDHAFTNARDALNVSOAWRWSLPYKVKSRLLQKACGDAMNLLGYR	360		
Db	301	HLQTVHNIITRGKMGDHAFTNARDALNVSOAWRWSLPYKVKSRLLQKACGDAMNLLGYR	360		
Oy	361	HVRSEQQRNLLDLLSTWTVPQI 386			

Db 361 HVRSEQEQRNLLDLLSTWTVPQIH 386

RESULT 2

US-09-190-911-1
; Sequence 1, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-190-911-1

Query Match 100.0%; Score 2038; DB 4; Length 386;

Best Local Similarity 100.0%; Pred. No. 3e-212;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
Db 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
Qy 61 LFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
Db 61 LFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
Qy 121 SSLFOWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVYLKEVR 180
Db 121 SSLFOWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVYLKEVR 180
Qy 181 FENLQSLYPLLDPSLNHLHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
Db 181 FENLQSLYPLLDPSLNHLHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
Qy 241 QPYVMQVTCQSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELP 300
Db 241 QPYVMQVTCQSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELP 300
Qy 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
Db 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
Qy 361 HVRSEQEQRNLLDLLSTWTVPQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPQIH 386

RESULT 3

US-09-786-240-11
; Sequence 11, Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. Iton
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZINZAI, Yalda

; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11

Query Match 95.0%; Score 1936; DB 4; Length 386;

Best Local Similarity 95.6%; Pred. No. 3.4e-201;

Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
Db 1 MLLPKMKLLLLFLVSQMAILALFFHMYSHNISLSMKQAQPERMHVLLSSWRSGSSFVGQ 60
Qy 61 LFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
Db 61 LFGQHPDVFYLMPEPAWHVMTFKQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
Qy 121 SSLFOWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVYLKEVR 180
Db 121 SSLFOWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQOPFEVVEKACRSYSHVYLKEVR 180
Qy 181 FENLQSLYPLLDPSLNHLHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
Db 181 FENLQSLYPLLDPSLNHLHVLVRDPRAVFRSRRRTKGLMIDSRIVMGQHEQKLKED 240
Qy 241 QPYVMQVTCQSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELP 300
Db 241 QPYVMQVTCQSQLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFELP 300
Qy 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
Db 301 HLOTWVHNITRGKMGDHAFTNARDALNVSOAWWSLPYEKVSRLQKACGDAMNLLGYR 360
Qy 361 HVRSEQEQRNLLDLLSTWTVPQIH 386
Db 361 HVRSEQEQRNLLDLLSTWTVPQIH 386

RESULT 4

US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Re-ji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: JP 10-177844
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2

```
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

Query Match      29.4%; Score 598.5; DB 3; Length 483;
Best Local Similarity 36.8%; Pred. No. 4.4e-56;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWSSGSSFFVGLFQHPDVFYLMPEPAHVMWTFKQSTAWMLHMAVRDLIRA 100
DB 116 KROLVYVFTTWRSRGSFFGELFNQNEVFVLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 175
QY 101 VFLCDMSVFDAYMEPPRRQS-----SLFQWENSRA LCSAPACDIIPODEIIIPRAHCRLL- 155
DB 176 LYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDRVCK 234
QY 156 -CSQOPFEVVEKACRSYSHVVLKEVFPFNLSLYPLLKDPKPSLNHLHIVHLVRDPAVRFSR 214
DB 235 KCPQRLARFECECKRYRTVVIKGVRFVDVAVLAPLLKDPALDLKVIHLVRDPAVASSR 294
QY 215 ERTKGLMIDSRIVM-----GQHEQKLKEDQ-----PYV---VMQVICO 252
DB 295 IRRSHGLRESLQVVRSDPRAHRMPEFLAAGHKLGAKEGMPADYHALGAMEVICNS 354
QY 253 QLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPLHLQTVWVHNITRG 312
DB 355 MAKTLQALQPPDWLQGHVLYVRYEDLVGDPVKTLLRVYDFVGLLVSPENEFQALNMTSG 414
QY 313 KGMGDHAFHTNARDALNVSAQWWSLPEYKVSRLQKACGDAMNLLGYRHSRSEQORNL 372
DB 415 SGSSSPFVVSARNATQANAWTALTFTQIKQVEEFQCPMAVLGYRVNSPEEVKDL 474
QY 373 LDLL 376
DB 475 KTL 478

RESULT 5
US-09-471-867-2
; Sequence 2, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; EARLIER FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

Query Match      29.4%; Score 598.5; DB 4; Length 483;
Best Local Similarity 36.8%; Pred. No. 4.4e-56;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWSSGSSFFVGLFQHPDVFYLMPEPAHVMWTFKQSTAWMLHMAVRDLIRA 100
DB 116 KROLVYVFTTWRSRGSFFGELFNQNEVFVLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 175
QY 101 VFLCDMSVFDAYMEPPRRQS-----SLFQWENSRA LCSAPACDIIPODEIIIPRAHCRLL- 155
DB 176 LYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDRVCK 234
QY 156 -CSQOPFEVVEKACRSYSHVVLKEVFPFNLSLYPLLKDPKPSLNHLHIVHLVRDPAVRFSR 214
DB 235 KCPQRLARFECECKRYRTVVIKGVRFVDVAVLAPLLKDPALDLKVIHLVRDPAVASSR 294
QY 215 ERTKGLMIDSRIVM-----GQHEQKLKEDQ-----PYV---VMQVICO 252
DB 295 IRRSHGLRESLQVVRSDPRAHRMPEFLAAGHKLGAKEGMPADYHALGAMEVICNS 354
QY 253 QLEIYKTIQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPLHLQTVWVHNITRG 312
DB 355 MAKTLQALQPPDWLQGHVLYVRYEDLVGDPVKTLLRVYDFVGLLVSPENEFQALNMTSG 414
QY 313 KGMGDHAFHTNARDALNVSAQWWSLPEYKVSRLQKACGDAMNLLGYRHSRSEQORNL 372
DB 415 SGSSSPFVVSARNATQANAWTALTFTQIKQVEEFQCPMAVLGYRVNSPEEVKDL 474
QY 373 LDLL 376
DB 475 KTL 478

RESULT 6
US-09-263-023-4
; Sequence 4, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-263-023-4

Query Match      28.8%; Score 587.5; DB 3; Length 484;
Best Local Similarity 36.0%; Pred. No. 6.8e-55;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWSSGSSFFVGLFQHPDVFYLMPEPAHVMWTFKQSTAWMLHMAVRDLIRA 100
DB 117 KRHWYVFTTWRSRGSFFGELFNQNEVFVLYEPVWHVWQKLYPGDAVSLQGAARDMLSA 176
QY 101 VFLCDMSVFDAYMEPPRRQS-----SLFQWENSRA LCSAPACDIIPODEIIIPRAHCRLL- 155
DB 177 LYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAY-RKEVVGVLVDRVCK 235
QY 156 -CSQOPFEVVEKACRSYSHVVLKEVFPFNLSLYPLLKDPKPSLNHLHIVHLVRDPAVRFSR 214
DB 236 KCPQRLARFECECKRYRTVVIKGVRFVDVAVLAPLLKDPALDLKVIHLVRDPAVASSR 295
QY 215 ERTKGLMIDSRIVM-----GQHEQKLKEDQ-----DQPIY---VMQVICO 252
```

```
Db 296 IRSRGLRESLQVVRSDPRAHRMPFLEAAGHKLGAKEGVGGPADYHALGAMEVICNS 355
QY 253 QLEIYKTTQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPLHLQVWVHNITRG 312
Db 356 MAKTLQALQPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDFVGLVSPMEQFALNMTSG 415
QY 313 KMGDGHAFHTNARDALNVSQAWRWSLPYKVSRLQKACGDAMNLLGYRHVSEOEQRNLL 372
Db 416 SGSSSKPFVVSARNATQANAWRTALTTCQIKQVEEFCYQPMVLGVERVNSPBEVKDLS 475
QY 373 LDLL 376
Db 476 KTL 479

RESULT 7
US-09-471-867-4
; Sequence 4, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kanragi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 28.8%; Score 587.5; DB 4; Length 484;
Best Local Similarity 36.0%; Pred. No. 6.8e-55;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLVLSWRSGSFFVQLFGQHPDVFYLMPEAWHVMWTFKOSTAWMLHMAVRDLIRA 100
Db 117 KHMVYVTTWRSGSFFGELFNQNPVEFFLYEPVWVWQKLYPCDAVSLQGAARDLSA 176
QY 101 VFLCDMSVFDAYMEGPRRQS----SLFWENSRLCSAPACDIIPDEIIPRAHCRLL- 155
Db 177 LYRCDLSVFLYSPAGSGRNLTILGIFGAATNKVCSPLCPAY-RKEVVLVDVDRUCK 235
QY 156 -CSQOPFEVVEKACRSYSHVVKVRFNPNLSQSLVPLKDPKSLNHLVHLVDRPRAVFRSR 214
Db 236 KCPQRLARFEECKRYRTLKIGVRVDFVAVLPLDRDPAKLVHLVDRPRAVASSR 295
QY 215 ERTKGLMIDSRIVN-----GQEQKLKKE---DQVY---VMQVICSQ 252
Db 296 IRSRGLRESLQVVRSDPRAHRMPFLEAAGHKLGAKEGVGGPADYHALGAMEVICNS 355
QY 253 QLEIYKTTQSLPKALQERYLLVRYEDLARAPVAQTSRMVYFVGLFPLHLQVWVHNITRG 312
Db 356 MAKTLQALQPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDFVGLVSPMEQFALNMTSG 415
QY 313 KMGDGHAFHTNARDALNVSQAWRWSLPYKVSRLQKACGDAMNLLGYRHVSEOEQRNLL 372
Db 416 SGSSSKPFVVSARNATQANAWRTALTTCQIKQVEEFCYQPMVLGVERVNSPBEVKDLS 475
```

```
QY 373 LDLL 376
Db 476 KTL 479

RESULT 8
US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match 26.9%; Score 549; DB 4; Length 411;
Best Local Similarity 34.3%; Pred. No. 7.9e-51;
Matches 139; Conservative 64; Mismatches 160; Indels 42; Gaps 10;

QY 6 KMKLLFLVS---QMAILALFFHMYSHNISSLSMKAQPERM-----HVLV 47
Db 6 KAVLLIALASIAIQVTAIRTTAKSFHTCPGLAEAGLAERLCEESPTFAYNLSRKHILI 65
QY 48 LSSWSGSSFFVQLFGQHPDVFYLMPEAWHVMWTFKOSTA---WMLHMAVRDLIR 99
Db 66 LATTREGSFFVQLFNQHLDFILFELPYHVQNTLIPFTQCKSPADRRVWLGASRDLLR 125
QY 100 AVFLCDMSVFDAYMEGPRRQS--LFWENSRLCSAPACDIIPDEIIPRAHCRLLC 156
Db 126 SLYDCDLYFLENYIKPPVNHHTDIFRRGASRVLCSPVCDPPGPAADLVLEEGDCVRKC 185
QY 157 SQOPFEVVEKACRSYSHVVKVRFNPNLSQSLVPLKDPKSLNHLVHLVDRPRAVFRSR 216
Db 186 GLNLTVAAEACRERSHVAIKTVRVFVNDLRALVEDPRLNKLVIQLVDRPGLIASRSE 245
QY 217 TKGDLMIDSRIVMGOHEQKLKKEQDQVYV---MQVICQSLIYKTTQSLPKALQERYL 272
Db 246 TFRDLYRLWRLWYGTGR-----KPYNLDVTLTTCEDFNSVSTGLMRPPLKGYM 298
QY 273 LYRYEDLARAPVAQTSRMVYFVGLFPLHLQVWVHNITRG-KMGDGHAFHTNARDALNVS 331
Db 299 LYRYEDLARPNMKKTEEYIGFLGIPLDGSHVARWICQNTRGDPTLGKHYGT-VRNAATA 357
QY 332 QAWRWSLPYKVSRLQKACGDAMNLLGYRHVSEOEQRNLLDLL 376
Db 358 EKWRFSLVDIVAFQAQACQVLAQLGYKIAASEELKNPVSIV 402

RESULT 9
US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASAKAZU
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
```

COUNTRY: ;
 ZIP: ;
 COMPUTER READABLE FORM: ;
 MEDIUM TYPE: Floppy disk ;
 COMPUTER: IBM PC compatible ;
 OPERATING SYSTEM: PC-DOS/MS-DOS ;
 SOFTWARE: PatentIn ;
 CURRENT APPLICATION DATA: ;
 APPLICATION NUMBER: US/08/655,878 ;
 FILING DATE: ;
 CLASSIFICATION: 435 ;
 PRIOR APPLICATION DATA: ;
 APPLICATION NUMBER: ;
 FILING DATE: ;
 ATTORNEY/AGENT INFORMATION: ;
 NAME: ;
 REGISTRATION NUMBER: ;
 REFERENCE/DOCKET NUMBER: ;
 TELEPHONE: ;
 TELEFAX: ;
 INFORMATION FOR SEQ ID NO: 2: ;
 SEQUENCE CHARACTERISTICS: ;
 LENGTH: 458 ;
 TYPE: amino acid ;
 TOPOLOGY: linear ;
 MOLECULE TYPE: protein ;
 US-08-655-878-2

Query Match 24.6%; Score 500.5; DB 2; Length 458;
 Best Local Similarity 33.2%; Pred. No. 1.7e-45;
 Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;
 QY 34 LNKAKQPERMHLVLSWSRSGSFVQQLFGQHPDVFLMEPAWHV--WMTFKQSTAWMLH 91
 DB 104 LGIAAEPRRHVLLMATTGTSGFGEFFNQGNIFLPEPLWHIERTVTPFGGANAVG 163
 QY 92 MAV--RDLIRAVFLCDMSVFDAYMEPPRRQ--SSLFQWNSRALCSAPACDIIQ-DEI 146
 DB 164 SALVYRDVQLLCCDLYLESFISPAPEHLTAALFRGSGSHSCEEPVC--TPSLKKV 221
 QY 147 IPRACR-LICSQPPEVVEKACRSYSHVVLKEVFFNLSQSLYPLKDPSLMLHLVHLVR 205
 DB 222 FEKYCKNRCRGPLNTLAAECRRKQHAKLVTRIRQLEFLQPLAEDPRDLRIQLVR 281
 QY 206 DPRVRSRERTKGLMIDSRIVMGQHEQKLK-----KEDQPYVMQVICO-S-QL 255
 DB 282 DPRVLVSR-----MV---AFSGKYESKWKWAAEGEAPLOEDE-VQSLRGNCESIRLS 330
 QY 256 IYKTIQSLKALQERYLLVRYEDLAPVPAQTSRMVYFVGLFELPHLOTWVHNITRGKM 315
 DB 331 AELGLRQ-PWLEGRVLYRYEDVAPAPLRKALEMYRFAGIHPTQVEWIRANTQAP-Q 388
 QY 316 GDHAFHTNARDALNVSQAWNSLPYKVSRLQKACGDAMNLLGYRHVRSEQEQRLNLLDL 375
 DB 389 DSNGIYSTOKNSSEQEQKWRFSIPFKLAQVVDACEPAMRLFGYKLASSAQELNRSLSL 448
 QY 376 L 376
 DB 449 L 449

RESULT 10
 US-08-899-514-2
 ; Sequence 2, Application US/08899514
 ; Patent No. 5910581
 ; GENERAL INFORMATION:
 ; APPLICANT: HABUCHI, OSAMI
 ; APPLICANT: FUKUTA, WASAKAZU
 ; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 ; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 ; TITLE OF INVENTION: FOR THE POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS: ;
 ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP ;
 STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR ;
 CITY: NEWPORT BEACH ;
 STATE: CALIFORNIA ;
 COUNTRY: US ;
 ZIP: 92660 ;
 COMPUTER READABLE FORM: ;
 MEDIUM TYPE: Floppy disk ;
 COMPUTER: IBM PC compatible ;
 OPERATING SYSTEM: PC-DOS/MS-DOS ;
 SOFTWARE: PatentIn ;
 CURRENT APPLICATION DATA: ;
 APPLICATION NUMBER: US/08/899,514 ;
 FILING DATE: ;
 CLASSIFICATION: 435 ;
 PRIOR APPLICATION DATA: ;
 APPLICATION NUMBER: ;
 FILING DATE: ;
 ATTORNEY/AGENT INFORMATION: ;
 NAME: DANIEL E. ALTMAN ;
 REGISTRATION NUMBER: 34,115 ;
 REFERENCE/DOCKET NUMBER: TOYAM21.001AUS ;
 TELECOMMUNICATION INFORMATION: ;
 TELEPHONE: 714 760 0404 ;
 TELEFAX: 714 760 9502 ;
 INFORMATION FOR SEQ ID NO: 2: ;
 SEQUENCE CHARACTERISTICS: ;
 LENGTH: 479 ;
 TYPE: amino acid ;
 TOPOLOGY: linear ;
 MOLECULE TYPE: protein ;
 US-08-899-514-2

Query Match 23.7%; Score 482; DB 2; Length 479;
 Best Local Similarity 32.4%; Pred. No. 1.8e-43;
 Matches 113; Conservative 71; Mismatches 141; Indels 24; Gaps 10;
 QY 42 RMHVLVLSWSRSGSFVQQLFGQHPDVFLMEPAWHVMT--FKQSTAWMLHMAV--RDL 97
 DB 132 RRHLLMATTGTSGFGEFFNQGNIFLPEPLWHIERTVTPFGGANAVG 191
 QY 98 IRVFLCDMSVFDAYMEPPRRQSSLFQWE--NSRALCSAPACDIIQDEIIPRAHCR-L 154
 DB 192 LKQLFLCDLVLEHFTPLPEDHLTQFMFRGSSRSICEDPVCTPFVK-KVFEKYHCKNR 250
 QY 155 LCSQPPPEVVEKACRSYSHVVLKEVFFNLSQSLYPLKDPSLMLHLVHLVDRPVRFR 214
 DB 251 RCGPLNVTLAAECRRKQHAKLVTRIRQLEFLQPLAEDPRDLRIQLVDRPVLASR 310
 QY 215 -----ERTKGLMIDSRIVMGQHEQKLKEDQPYVMQVICO-S-OLEIYKTIQSLPKAL 267
 DB 311 MVAFAQYKTKWKWLDDEGGDGLREEVQR-----LRGNCESIRLSAELGRQ-PAWL 362
 QY 268 QERYLLVRYEDLAPVPAQTSRMVYFVGLFELPHLOTWVHNITRGKMGDHAFTNARDA 327
 DB 363 RGRVLYRYEDVARGPLQKAREMYFPAGIPLTPQVEDWIKNTQAAHDSGGIYSTQ-KNS 421
 QY 328 LNVSQAWNSLPYKVSRLQKACGDAMNLLGYRHVRSEQEQRLNLLDL 376
 DB 422 SEQEKWRFSMPFKLAQVVDACEPAMRLFGYKLARDAAALNRSVSL 470

RESULT 11
 US-08-811-469-6
 ; Sequence 6, Application US/09811469
 ; Patent No. 6551809
 ; GENERAL INFORMATION:
 ; APPLICANT: YAN, Chunhua et al.
 ; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
 ; TITLE OF INVENTION: AND USES THEREOF
 ; FILE REFERENCE: C1001171

